(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 October 2002 (03.10.2002)

PCT

(10) International Publication Number
WO 02/077237 A2

(51) International Patent Classification⁷: C12N 15/12, C07K 14/05, C12N 5/10, C12Q 1/68, C07K 16/18, A61K 38/17, 39/395, A01K 67/027, G01N 33/50, 33/53

(21) International Application Number: PCT/US02/03657

(22) International Filing Date: 8 February 2002 (08.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/267,892	9 February 2001 (09.02.2001)	US
60/271,168	23 February 2001 (23.02.2001)	US
60/272,890	2 March 2001 (02.03.2001)	US
60/276,860	16 March 2001 (16.03.2001)	US
60/278,255	23 March 2001 (23.03.2001)	US
60/280,538	30 March 2001 (30.03.2001)	US
60/351,359	25 January 2002 (25.01.2002)	US

(71) Applicant *(for all designated States except US)*: INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee St. #5, San Francisco, CA 94107 (US). BRUNS, Christopher, M. [US/US]; 2255 Showers Drive #264, Mountain View, CA 94040 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polten Place Way, San Jose, CA 95121 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). RAUMANN, Brigitte, E. [US/US]; 5801 South Dorchester Avenue #3B, Chicago, IL 60637 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). LAL, Preeti, G. [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). ARVIZU, Chandra [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). CHEN, Hsei-Mei [US/US]; 2425 Lyle Court, Santa Leandro, CA 94578 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 02/077237 A2

TRANSPORTERS AND ION CHANNELS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including K^+, NH_4^+, P_i, SO_4^{2-}, sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous Na^+/K^+ ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging techniques

and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H^+-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H^+-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na^+-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective

transporters for organic cations and organic anions in organs including the kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) Am. J. Physiol. 264:C761-C782; Price, N.T. et al. (1998) Biochem. J. 329:321-328; and Martinelle, K. and I. Haggstrom (1993) J. Biotechnol. 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) Meth. Enzymol. 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) J. Med. Genet. 23:99-106).

P-type ATPases comprise a class of cation-transporting transmembrane proteins. They are integral membrane proteins which use an aspartyl phosphate intermediate to move cations across a membrane. Features of P-type ATPases include: (i) a cation channel; (ii) a stalk, formed by

extensions of the transmembrane α-helices into the cytoplasm; (iii) an ATP binding domain; (iv) a phosphorylated aspartic acid; (v) an adjacent transduction domain; (vi) a phosphatase domain, which removes the phosphate from the aspartic acid as part of the reaction cycle; and (vii) six or more transmembrane domains. Included in this class are heavy metal-transporting ATPases as well as aminophospholipid transporters.

The transport of phosphatidylserine and phosphatidylethanolamine by aminophospholipid translocase results in the movement of these molecules from one side of a bilayer to another. This transport is conducted by a newly identified subfamily of P-type ATPases which are proposed to be amphipath transporters. Amphipath transporters move molecules having both a hydrophilic and a hydrophobic region. As many as seventeen different genes belong to this P-type ATPases subfamily, being grouped into several distinct classes and subclasses (Halleck, M.S. et al., (1999) Physiol. Genomics 1:139-150; Vulpe,C. et al., (1993) Nat. Genet. 3:7-13).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process. Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

The lipocalin superfamily constitutes a phylogenetically conserved group of more than forty proteins that function as extracellular ligand-binding proteins which bind and transport small hydrophobic molecules. Members of this family function as carriers of retinoids, odorants, chromophores, pheromones, allergens, and sterols, and in a variety of processes including nutrient transport, cell growth regulation, immune response, and prostaglandin synthesis. A subset of these proteins may be multifunctional, serving as either a biosynthetic enzyme or as a specific enzyme inhibitor. (Tanaka, T. et al. (1997) J. Biol. Chem. 272:15789-15795; and van't Hof, W. et al. (1997) J. Biol. Chem. 272:1837-1841.)

Members of the lipocalin family display unusually low levels of overall sequence conservation. Pairwise sequence identity often falls below 20%. Sequence similarity between family members is limited to conserved cysteines which form disulfide bonds and three motifs which form a juxtaposed cluster that functions as a target cell recognition site. The lipocalins share an eight stranded, anti-parallel beta-sheet which folds back on itself to form a continuously hydrogen-bonded beta-barrel. The pocket formed by the barrel functions as an internal ligand binding site. Seven loops (L1 to L7)

form short beta-hairpins, except loop L1 which is a large omega loop that forms a lid to partially close the internal ligand-binding site (Flower (1996) Biochem. J. 318:1-14).

Lipocalins are important transport molecules. Each lipocalin associates with a particular ligand and delivers that ligand to appropriate target sites within the organism. Retinol-binding protein (RBP), one of the best characterized lipocalins, transports retinol from stores within the liver to target tissues. Apolipoprotein D (apo D), a component of high density lipoproteins (HDLs) and low density lipoproteins (LDLs), functions in the targeted collection and delivery of cholesterol throughout the body. Lipocalins are also involved in cell regulatory processes. Apo D, which is identical to gross-cystic-disease-fluid protein (GCDFP)-24, is a progesterone/pregnenolone-binding protein expressed at high levels in breast cyst fluid. Secretion of apo D in certain human breast cancer cell lines is accompanied by reduced cell proliferation and progression of cells to a more differentiated phenotype. Similarly, apo D and another lipocalin, α_1-acid glycoprotein (AGP), are involved in nerve cell regeneration. AGP is also involved in anti-inflammatory and immunosuppressive activities. AGP is one of the positive acute-phase proteins (APP); circulating levels of AGP increase in response to stress and inflammatory stimulation. AGP accumulates at sites of inflammation where it inhibits platelet and neutrophil activation and inhibits phagocytosis. The immunomodulatory properties of AGP are due to glycosylation. AGP is 40% carbohydrate, making it unusually acidic and soluble. The glycosylation pattern of AGP changes during acute-phase response, and deglycosylated AGP has no immunosuppressive activity (Flower (1994) FEBS Lett. 354:7-11; Flower (1996) supra).

The lipocalin superfamily also includes several animal allergens, including the mouse major urinary protein (mMUP), the rat α-2-microgloobulin (rA2U), the bovine β-lactoglobulin (βlg), the cockroach allergen (Bla g4), bovine dander allergen (Bos d2), and the major horse allergen, designated *Equus caballus* allergen 1 (Equ c1). Equ c1 is a powerful allergen responsible for about 80% of anti-horse IgE antibody response in patients who are chronically exposed to horse allergens. It appears that lipocalins may contain a common structure that is able to induce the IgE response (Gregoire, C. et al., (1996) J. Biol. Chem. 271:32951-32959).

Lipocalins are used as diagnostic and prognostic markers in a variety of disease states. The plasma level of AGP is monitored during pregnancy and in diagnosis and prognosis of conditions including cancer chemotherapy, renal disfunction, myocardial infarction, arthritis, and multiple sclerosis. RBP is used clinically as a marker of tubular reabsorption in the kidney, and apo D is a marker in gross cystic breast disease (Flower (1996) supra). Additionally, the use of lipocalin animal allergens may help in the diagnosis of allergic reactions to horses (Gregoire supra), pigs, cockroaches, mice and rats.

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and

charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) *275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int. Med. 245:637-642).

Ion Channels

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient. These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including Na^+-K^+ ATPase, Ca^{2+}-ATPase, and H^+-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na^+ and Ca^{2+} are low and cytosolic concentration of K^+ is high. The vacuolar (V) class of ion transporters includes H^+ pumps on intracellular organelles,

such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H^+ pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P_i).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). P-type ATPases use an aspartyl phosphate intermediate to move cations across a membrane. Features of P-type ATPases include: (i) a cation channel; (ii) a stalk, formed by extensions of the transmembrane α-helices into the cytoplasm; (iii) an ATP binding domain; (iv) a phosphorylated aspartic acid; (v) an adjacent transduction domain; (vi) a phosphatase domain, which removes the phosphate from the aspartic acid as part of the reaction cycle; and (vii) six or more transmembrane domains. Included in this class are heavy metal-transporting ATPases as well as aminophospholipid transporters. The FIC1 gene encodes a P-type ATPase that is mutated in two forms of hereditary cholestasis. The protein product of FIC1 is likely to play an essential role in bile acid circulation in the liver (Bull, L.N. et al. (1998) Nat. Genet. 18:219-224). The V-ATPases are composed of two functional domains: the V_1 domain, a peripheral complex responsible for ATP hydrolysis; and the V_0 domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase F_0 domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V_0 domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgac, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na^+ down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca^{2+} out of the cell with transport of Na^+ into the cell (antiport).

Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction,

fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na^+, K^+, Ca^{2+}, and Cl^- channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g., acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both Ca^{2+} and Na^+ (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the Na^+ and Ca^{2+} subfamilies, this domain is repeated four times, while in the K^+ channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K^+ channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na^+ and K^+ channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na^+ and K^+ ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na^+ channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated Na^+ channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's

N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na^+ channels are heterotrimeric complexes composed of a 260 kDa pore-forming α subunit that associates with two smaller auxiliary subunits, β1 and β2. The β2 subunit is a integral membrane glycoprotein that contains an extracellular Ig domain, and its association with α and β1 subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na^+ channels include the members of the amiloride-sensitive Na^+ channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na^+ channel (ENaC) involved in Na^+ reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized H^+-gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na^+-permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from C. elegans. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglen, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K^+ channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca^{2+} and cAMP. In non-excitable tissue, K^+ channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K^+ channels are responsible for setting the resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na^+-K^+ pump and ion channels that provide the redistribution of Na^+, K^+, and Cl^-. The pump actively transports Na^+ out of the cell and K^+ into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow K^+ and Cl^- to flow by passive diffusion. Because of the high negative charge within the cytosol, Cl^- flows out of the cell. The flow of K^+ is balanced by an electromotive force pulling K^+ into the cell, and a K^+ concentration gradient pushing K^+ out of the cell. Thus, the resting membrane potential is primarily regulated by K^+ flow (Salkoff, L. and T. Jegla (1995) Neuron 15:489-492).

Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic β subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated K^+ channels as well as the delayed rectifier type channels such as the human ether-a-go-go related gene (HERG) associated with long QT, a cardiac dysrythmia syndrome (Curran, M.E. (1998) Curr. Opin. Biotechnol. 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) Curr. Opin. Chem. Biol. 3:448-458).

A second superfamily of K^+ channels is composed of the inward rectifying channels (Kir). Kir channels have the property of preferentially conducting K^+ currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K^+ channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) Curr. Opin. Neurobiol. 5:268-277; Curran, supra).

The recently recognized TWIK K^+ channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential in a large set of cell types (Duprat, F. et al. (1997) EMBO J 16:5464-5471).

The voltage-gated Ca^{2+} channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca^{2+} channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca^{2+} channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The α_1 subunit forms the membrane pore and voltage sensor, while the $\alpha_2\delta$ and β subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel. These subunits are encoded by at least six α_1, one $\alpha_2\delta$, and four β genes. A fourth subunit, γ, has been identified in skeletal muscle (Walker, D. et al. (1998) J. Biol. Chem. 273:2361-2367; McCleskey, E.W. (1994) Curr. Opin. Neurobiol. 4:304-312).

The high-voltage-activated Ca^{2+} channels that have been characterized biochemically include complexes of a pore-forming alpha1 subunit of approximately 190-250 kDa; a transmembrane

complex of alpha2 and delta subunits; an intracellular beta subunit; and in some cases a transmembrane gamma subunit. A variety of alpha1 subunits, alpha2delta complexes, beta subunits, and gamma subunits are known. The Cav1 family of alpha1 subunits conduct L-type Ca^{2+} currents, which initiate muscle contraction, endocrine secretion, and gene transcription, and are regulated primarily by second messenger-activated protein phosphorylation pathways. The Cav2 family of alpha1 subunits conduct N-type, P/Q-type, and R-type Ca^{2+} currents, which initiate rapid synaptic transmission and are regulated primarily by direct interaction with G proteins and SNARE proteins and secondarily by protein phosphorylation. The Cav3 family of alpha1 subunits conduct T-type Ca^{2+} currents, which are activated and inactivated more rapidly and at more negative membrane potentials than other Ca^{2+} current types. The distinct structures and patterns of regulation of these three families of Ca^{2+} channels provide an array of Ca^{2+} entry pathways in response to changes in membrane potential and a range of possibilities for regulation of Ca^{2+} entry by second messenger pathways and interacting proteins (Catterall, W.A. (2000) Annu. Rev. Cell Dev. Biol. 16:521-555).

The alpha-2 subunit of the voltage-gated Ca^{2+}-channel may include one or more Cache domains. An extracellular Cache domain may be fused to an intracellular catalytic domain, such as the histidine kinase, PP2C phosphatase, GGDEF (a predicted diguanylate cyclase), HD-GYP (a predicted phosphodiesterase) or adenylyl cyclase domain, or to a noncatalytic domain, like the methyl-accepting, DNA-binding winged helix-turn-helix, GAF, PAS or HAMP (a domain found in istidine kinases, denylyl cyclases, ethyl-binding proteins and phosphatases). Small molecules are bound via the Cache domain and this signal is converted into diverse outputs depending on the intracellular domains (Anantharaman, V. and Aravind, L.(2000) Trends Biochem. Sci. 25:535-537).

The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the Ca^{2+} influx into cells to resupply Ca^{2+} stores depleted by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from Drosophila and have similarity to voltage gated Ca^{2+} channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCE channels (Zhu, X. et al. (1996) Cell 85:661-671; Boulay, G. et al. (1997) J. Biol. Chem. 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, whose expression in melanoma cells is inversely correlated with melanoma aggressiveness in vivo. The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) J. Clin. Oncol. 19:568-576).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and

organelle pH. In secretory epithelial cells, Cl^- enters the cell across a basolateral membrane through an Na^+, K^+/Cl^- cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl^- from the apical surface, in response to hormonal stimulation, leads to flow of Na^+ and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) J. Exp. Biol. 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) Curr. Opin. Neurobiol. 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of Na^+ and produces the initial localized depolarization that activates the voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as γ-aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, supra). Amino acids in the second transmembrane domain appear to be important in determining channel permeation and selectivity (Sather, W.A. et al. (1994) Curr. Opin. Neurobiol. 4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated K^+ channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens K^+ channels to modulate the magnitude of the action potential (Ishi et al., supra). The large conductance (BK) channel has been purified from brain and its subunit composition

determined. The α subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated K^+ channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The β subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, supra; Vergara, C. et al. (1998) Curr. Opin. Neurobiol. 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na^+ channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for Ca^{2+} entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an α subunit which can form functional homomeric channels, and a β subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K^+ channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) Curr. Opin. Neurobiol. 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the Gβγ subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) Curr. Opin. Cell. Biol. 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Bredt (1998) Cell 93:495-498).

Disease Correlation

The etiology of numerous human diseases and disorders can be attributed to defects in the transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff,

W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G.M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Engl. J. Med. 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrythmia disorders such as the long QT syndromes and idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan (1998) Proc. Natl. Acad. Sci. USA 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) Curr. Opin. Neurology 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) Curr. Opin. Neurobiol. 9:274-280; Cooper, supra).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) Adv. Pharmacol. 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na^+ channels have been useful in the treatment of neuropathic pain (Eglen, supra).

Ion channels in the immune system have recently been suggested as targets for immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

Senescence

Most normal eukaryotic cells, after a certain number of divisions, enter a state of senescence in which cells remain viable and metabolically active but no longer replicate. A number of phenotypic

changes such as increased cell size and pH-dependent beta-galactosidase activity, and molecular changes such as the upregulation of particular genes, occur in senescent cells (Shelton (1999) Current Biology 9:939-945). When senescent cells are exposed to mitogens, a number of genes are upregulated, but the cells do not proliferate. Evidence indicates that senescent cells accumulate with age in vivo, contributing to the aging of an organism. In addition, senescence suppresses tumorigenesis, and many genes necessary for senescence also function as tumor suppressor genes, such as p53 and the retinoblastoma susceptibility gene. Most tumors contain cells that have surpassed their replicative limit, i.e. they are immortalized. Many oncogenes immortalize cells as a first step toward tumor formation.

A variety of challenges, such as oxidative stress, radiation, activated oncoproteins, and cell cycle inhibitors, induce a senescent phenotype, indicating that senescence is influenced by a number of proliferative and anti-proliferative signals (Shelton supra). Senescence is correlated with the progressive shortening of telomeres that occurs with each cell division. Expression of the catalytic component of telomerase in cells prevents telomere shortening and immortalizes cells such as fibroblasts and epithelial cells, but not other types of cells, such as CD8+ T cells (Migliaccio et al. (2000) J. Immunol. 165:4978-4984). Thus, senescence is controlled by telomere shortening as well as other mechanisms depending on the type of cell.

A number of genes that are differentially expressed between senescent and presenescent cells have been identified as part of ongoing studies to understand the role of senescence in aging and tumorigenesis. Most senescent cells are growth arrested in the G1 stage of the cell cycle. While expression of many cell cycle genes is similar in senescent and presenescent cells (Cristofalo (1992) Ann. N. Y. Acad. Sci. 663:187-194), expression of others genes such as cyclin-dependent kinases p21 and p16, which inhibit proliferation, and cyclins D1 and E is elevated in senescent cells. Other genes that are not directly involved in the cell cycle are also upregulated such as extracellular matrix proteins fibronectin, procollagen, and osteonectin; and proteases such as collagenase, stromelysin, and cathepsin B (Chen (2000) Ann. N.Y. Acad. Sci. 908:111-125). Genes underexpressed in senescent cells include those that encode heat shock proteins, c-fos, and cdc-2 (Chen supra).

P-glycoprotein is a member of the ABC transporter family that is expressed on cells of the immune system and plays a role in the secretion of cytokines and cytotoxic molecules. P-glycoprotein expression and function were found to be increased in aging lymphocytes. These differences may play a role in the changes in immune response, including increased frequency of infections and autoimmune phenomena, associated with human aging (Aggrawal, S. et al. (1997) J. Clin. Immunol. 17:448-454).

The discovery of new transporters and ion channels, and the polynucleotides encoding them,

satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2," "TRICH-3," "TRICH-4," "TRICH-5," "TRICH-6," "TRICH-7," "TRICH-8," "TRICH-9," "TRICH-10," "TRICH-11," "TRICH-12," "TRICH-13," "TRICH-14," "TRICH-15," "TRICH-16," "TRICH-17," "TRICH-18," "TRICH-19," and "TRICH-20." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-20. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:21-40.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the

amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so

forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological

or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a

specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr

Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:21-40 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:21-40, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:21-40 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:21-40 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:21-40 and the region of SEQ ID NO:21-40 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-20 is encoded by a fragment of SEQ ID NO:21-40. A fragment of SEQ ID NO:1-20 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-20. For example, a fragment of SEQ ID NO:1-20 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-20. The precise length of a fragment of SEQ ID NO:1-20 and the region of SEQ ID NO:1-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as

follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid

sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which

comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection

programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid

sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The

presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte

polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transporters and ion channels. For example, SEQ ID NO:3 is 85% identical, from residue M27 to residue N989, to rabbit anion exchanger 4a (GenBank ID g11611537) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a HCO^{3-} transporter family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is an anion exchanger.

In another example, SEQ ID NO:6 is 47% identical, from residue S7 to residue E350, to hamster Na+ dependent ileal bile acid transporter (GenBank ID g455033) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.7e-88, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a sodium bile acid symporter family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses using the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:6 is a sodium/bile acid symporter.

In another example, SEQ ID NO:9 is 68% identical, from residue E6 to residue I349, to mouse Ac39/physophilin, a subunit of the vacuolar ATPase (GenBank ID g1226235) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.2e-130, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains an ATP synthase (C/AC39) subunit domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses using the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID

NO:9 is a vacuolar ATPase subunit.

In another example, SEQ ID NO:10 is 83% identical, from residue M154 to residue R591, to murine melastatin (GenBank ID g3047272) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.6e-200, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains a transient receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, analysis provide further corroborative evidence that SEQ ID NO:10 is a calcium ion channel (note that melastatin has homology to members of the "transient receptor" family of "calcium channels").

In another example, SEQ ID NO:12 is 51% identical, from residue G761 to residue E1326, to rat multidrug resistance protein MRP5 (GenBank ID g6682827) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.5e-236, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains two ABC transporter transmembrane regions and two ABC transporter domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:12 is an ABC transporter.

For example, SEQ ID NO:18 is 76% identical, from residue M1 to residue D597, to rat renal osmotic stress-induced Na-Cl organic solute cotransporter (GenBank ID g531469) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2e-260, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:18 also contains a sodium:neurotransmitter symporter family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:18 is a sodium dependent organic solute transporter. SEQ ID NO:1-2, SEQ ID NO:4-5, SEQ ID NO:7-8, SEQ ID NO:11, SEQ ID NO:13-17 and SEQ ID NO:19-20 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-20 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide

consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:21-40 or that distinguish between SEQ ID NO:21-40 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis

methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:21-40, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a

polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:21-40. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding TRICH. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding TRICH, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding TRICH. For example, a polynucleotide comprising a sequence of SEQ ID NO:40 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:29. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences

include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:21-40 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising

a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

The nucleotide sequences of the present invention can be engineered using methods generally

known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TRICH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New

York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc.

Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase

such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that

TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in

the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a

convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. In addition, examples of tissues expressing TRICH are primary human breast epithelial cells and also can be found in Table 6. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, glycogen storage disease, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, pseudohypoaldosteronism type 1, Liddle's syndrome, cystinuria, iminoglycinuria, Hartup disease, Fanconi disease, and Bartter syndrome; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy,

retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, hemiplegic migraine, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, acid maltase deficiency (AMD, also known as Pompe's disease), generalized myotonia, and myotonia congenita; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial,

fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological and cell proliferative disorders described above. In one aspect, an antibody which specifically binds TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda,

S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies for TRICH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined

immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas

(Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While

alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA. by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA

sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is

detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the

case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the

subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess

expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:21-40 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for DNAs encoding TRICH include the cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility,

pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, glycogen storage disease, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, pseudohypoaldosteronism type 1, Liddle's syndrome, cystinuria, iminoglycinuria, Hartup disease, Fanconi disease, and Bartter syndrome; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, hemiplegic migraine, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, acid maltase deficiency (AMD, also known as Pompe's disease), generalized myotonia, and myotonia congenita; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy

(APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or

amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as

DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested

compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial

sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad.

Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant

invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/267,892, U.S. Ser. No. 60/271,168, U.S. Ser. No. 60/272,890, U.S. Ser. No. 60/276,860, U.S. Ser. No. 60/278,255, U.S. Ser. No. 60/280,538 and U.S. Ser. No. [Attorney Docket No. PF-1366, filed January 25, 2002] are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid,

QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic,

I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:21-40. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative transporters and ion channels were initially identified by running the Genscan gene

identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals

thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of TRICH Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:21-40 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:21-40 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between

chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C,

3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in TRICH Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:21-40 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:21-40 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based

hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is

reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

For SEQ ID NO:36, for example, HMECs, which are a primary human breast epithelial cell line isolated from a normal donor, were grown in Mammary Epithelial Cell Growth Medium (Clonetics, Walkersville MD) supplemented with 10 ng/ml human recombinant epidermal growth factor, 5 mg/ml insulin, 0.5 mg/ml hydrocortisone, 50 mg/ml gentamicin, 50 ng/ml amphotericin-B, and 0.5 mg/ml bovine pituitary extract. Cells were grown to 70-80% confluence prior to harvesting. About 1 x 10^7 cells were harvested at passage 8 (progenitor cells), passages 10 and 12 (progressively senescent cells), passage 14 (presenescent cells), and passage 15 (senescent cells). In this manner, it was demonstrated that the expression in senescent cells of component 2812176 of SEQ ID NO:36 is increased by a factor of at least 2.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a

cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

XIII. Expression of TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP;

Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of TRICH Specific Antibodies

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate,

blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring TRICH Using Specific Antibodies

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

XVII. Identification of Molecules Which Interact with TRICH

TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, molecules interacting with TRICH are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Identification of Molecules Which Interact with TRICH

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as Gβγ proteins (Reimann, supra) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, supra). TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are

assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH ion channel activity using the assays described in section XVIII.

XVIII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as ß-galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and ß-galactosidase.

Transformed cells expressing ß-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or ß-galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel,

and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing Xenopus laevis oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., supra; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate Xenopus oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18°C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or Ca^{+2} (in the form of $CaCl_2$), where appropriate. Electrode resistance is set at 2-5 MΩ and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 1mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH , 50µg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with ^3H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate. In particular, test substrates include glucose and other sugars for TRICH-1, aminophospholipids for TRICH-2, HCO^{3-} for TRICH-3, sulfate and other anions for TRICH-4, nucleotides for TRICH-5, Na^+ and bile acids for TRICH-6, TRICH-8, cationic amino acids for TRICH-11, amino acids for TRICH-7, protons for TRICH-9, drugs for TRICH-12, bile acids for TRICH-13 and TRICH-17, nucleosides for TRICH-15, drugs and other xenobiotics for TRICH-16, and neurotransmitters or organic osmolytes for TRICH-18.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP-[γ-^{32}P], separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered ^{32}P using a scintillation counter. The reaction mixture contains ATP-[γ-^{32}P] and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an

aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of ^{32}P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

Lipocalin activity of TRICH is measured by ligand fluorescence enhancement spectrofluorometry (Lin et al. (1997) Molecular Vision 3:17). Examples of ligands include retinol (Sigma, St. Louis MO) and 16-anthryloxy-palmitic acid (16-AP) (Molecular Probes Inc., Eugene OR). Ligand is dissolved in 100% ethanol and its concentration is estimated using known extinction coefficents (retinol: 46,000 A/M/cm at 325 nm; 16-AP: 8,200 A/M/cm at 361 nm). A 700 μl aliquot of 1 μM TRICH in 10 mM Tris (pH 7.5), 2 mM EDTA, and 500 mM NaCl is placed in a 1 cm path length quartz cuvette and 1 μl aliquots of ligand solution are added. Fluorescence is measured 100 seconds after each addition until readings are stable. Change in fluorescence per unit change in ligand concentration is proportional to TRICH activity.

In particular, the activity of TRICH-10 is measured as Ca^{2+} conductance, the activity of TRICH-14 is measured as K^+ conductance and the activity of TRICH-19 is measured as calcium-activated K+ conductance.

XIX. Identification of TRICH Agonists and Antagonists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVIII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the Ca^{2+} indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl^- indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonyl dyes such as $DiBAC_4$ (Molecular Probes). $DiBAC_4$ equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of $DiBAC_4$ entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
6911460	1	6911460CD1	21	6911460CB1
55138203	2	55138203CD1	22	55138203CB1
7478871	3	7478871CD1	23	7478871CB1
7483601	4	7483601CD1	24	7483601CB1
7487851	5	7487851CD1	25	7487851CB1
7472881	6	7472881CD1	26	7472881CB1
7612560	7	7612560CD1	27	7612560CB1
2880370	8	2880370CD1	28	2880370CB1
6267489	9	6267489CD1	29	6267489CB1
7484777	10	7484777CD1	30	7484777CB1
2493969	11	2493969CD1	31	2493969CB1
3244593	12	3244593CD1	32	3244593CB1
4921451	13	4921451CD1	33	4921451CB1
5547443	14	5547443CD1	34	5547443CB1
56008413	15	56008413CD1	35	56008413CB1
6127911	16	6127911CD1	36	6127911CB1
6427133	17	6427133CD1	37	6427133CB1
7472932	18	7472932CD1	38	7472932CB1
8463147	19	8463147CD1	39	8463147CB1
7506408	20	7506408CD1	40	7506408CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	6911460CD1	g145321	2.30E-65	[Escherichia coli] arabinose-proton symporter Maiden, M. C. J. et al. (1988) J. Biol. Chem. 263:8003-8010
2	55138203CD1	g4972583	0	[Homo sapiens] ATPase II Mouro, I. et al. (1999) Biochem. Biophys. Res. Commun. 257:333-339
3	7478871CD1	g11611537	0	[Oryctolagus cuniculus] anion exchanger 4a Tsuganezawa, H. et al. (2000) J. Biol. Chem. 276:8180-8189
4	7483601CD1	g8050590	6.30E-258	[Meriones unguiculatus] prestin Zheng, J. et al. (2000) Nature 405:149-155
5	7487851CD1	g1002424	2.40E-249	[Mus musculus] YSPL-1 (yolk sac permease-like molecule 1) form 1 Guimaraes, M. J. et al. (1995) Development 121:3335-3346
6	7472881CD1	g455033	3.70E-88	[Cricetulus griseus] Na+ dependent ileal bile acid transporter Wong, M.H. et al. (1994) J. Biol. Chem. 269:1340-1347
7	7612560CD1	g14571904	0	[Rattus norvegicus] lysosomal amino acid transporter 1 Sagne, C. et al. (2001) Proc. Natl. Acad. Sci. U.S.A. 98:7206-7211
8	2880370CD1	g455033	3.10E-36	[Cricetulus griseus] Na+ dependent ileal bile acid transporter Wong, M.H. et al. (1994) supra
9	6267489CD1	g1226235	3.20E-130	[Mus musculus] Ac39/physophilin Carrion-Vazquez, M. et al. (1998) Eur. J. Neurosci. 10:1153-66
10	7484777CD1	g3243075	0	[Homo sapiens] melastatin 1 Hunter,J.J. et al. (1998) Genomics 54:116-123 Duncan,L.M. et al. (2001) J. Clin. Oncol. 19:568-576
11	2493969CD1	g1589917	3.20E-137	[Rattus norvegicus] cationic amino acid transporter-1 Aulak, K.S. et al. (1996) J. Biol. Chem. 271:29799-29806
12	3244593CD1	g6682827	3.50E-236	[Rattus norvegicus] multidrug resistance protein (MRP5)
13	4921451CD1	g3628757	2.70E-257	[Homo sapiens] FIC1 Bull, L.N. et al. (1998) Cholestasis. Nat. Genet. 18:219-224

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
15	5600813CD1	g8698687	5.10E-29	[Mus musculus] equilibrative nitrobenzylthioinosine-insensitive nucleoside transporter ENT2 Kiss, A. et al. (2000) Biochem. J. 352:363-372
16	6127911CD1	g17223626	0	[Homo sapiens] ATP-binding cassette A10
17	6427133CD1	g3628757	0	[Homo sapiens] FIC1 Bull, L.N. et al. (1998) Cholestasis. Nat. Genet. 18:219-224
18	7472932CD1	g531469	1.20E-260	[Rattus norvegicus] renal osmotic stress-induced Na-Cl organic solute cotransporter Wasserman, J. C. et al. (1994) Am. J. Physiol. 267:F688-94
19	8463147CD1	g3978472	0	[Rattus norvegicus] potassium channel subunit Joiner, W. J. et al. (1998) Nat. Neurosci. 1:462-469
20	7506408CD1	g3955100	9.40E-71	[Mus musculus] vacuolar adenosine triphosphatase subunit D
		586887|Atp6d	7.90E-72	[Mus musculus] [Regulatory subunit; Active transporter, primary; Hydrolase; Transporter; ATPase] [Plasma membrane] Vacuolar H+-ATPase proton pump subunit D
		340040|ATP6D	7.10E-71	[Homo sapiens] [Regulatory subunit; Active transporter, primary; Hydrolase; Transporter; ATPase] [Plasma membrane] Vacuolar H+-ATPase proton pump (subunit D), an accessory subunit in the peripheral catalytic V1 complex, may be involved in coupling ATP hydrolysis (V1 complex) and proton transport (V0 complex) Agarwal, A. K. and White, P. C. (2000) Biochem. Biophys. Res. Commun. 279:543-547

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	6911460CD1	617	S75 S169 S220 S256 S264 S385 S443 T18 T246 T403 T520	N371 N383 N396 N401	Sugar (and other) transporter domain: S43-L564	HMMER_PFAM
					Transmembrane Domains: E80-R106, A109-S129, I134-Y154, V168-A188, H194-M214, N274-Y300, A316-D339, G342-M370, A458-L485, G509-M537 N-terminus is non-cytosolic	TMAP
					Sugar transport proteins BL00216: G51-S62, L133-A182	BLIMPS_BLOCKS
					Sugar transport proteins signatures: L119-I184	PROFILESCAN
					Sugar transporter signature PR00171: G51-I61, I134-V153, L465-V486, S488-M500	BLIMPS_PRINTS
					Glucose transporter signature PR00172: I279-Y300, S317-V338, L524-F544, L465-S488, R498-L516, W529-I549	BLIMPS_PRINTS
					SUGAR TRANSPORT PROTEINS DM00135|P09830|101-452: L119-G362	BLAST_DOMO
					Sugar transport proteins signature 1: G97-S113	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	55138203CD1	1193	S32 S45 S54 S58 S202 S215 S245 S317 S353 S437 S472 S491 S534 S580 S586 S593 S644 S727 S796 S848 S943 S1131 S1167 S1175 T14 T85 T125 T164 T299 T454 T486 T552 T614 T621 T686 T758 T777 T1108 T1133 T1185 Y530 Y608 Y617 Y1031	N36 N308 N857	E1-E2 ATPase domain: K161-S204	HMMER_PFAM
					Transmembrane Domains: R103-S123 T130-I150 E320-W348 N368-K396 C891-F911 C921-E941 V969-G995 G1026-Y1054 V1079-T1104 N-terminus is non-cytosolic	TMAP
					E1-E2 ATPases phosphorylation site signature BL00154:G183-L200, V432-F450, D690-L730, T825-S848	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: A418-P466	PROFILESCAN
					P-type cation-transporting atpase superfamily signature PR00119: E213-Q227, F436-F450, A706-D716, I828-I847	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2 (cont.)					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATPBINDING PROTEIN PROBABLE CALCIUMTRANSPORTING CALCIUM TRANSPORT PD004657: S862-R1103 PD004932: R34-P133	BLAST_PRODOM
					CHROMAFFIN GRANULE ATPASE II HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATPBINDING HOMOLOG PD038238: T1104-W1193 PD030421: K732-I801	BLAST_PRODOM
					do ATPASE; CALCIUM; TRANSPORTING; DM02405|P39524|236-1049: L116-N926	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): A770-T777, G1124-S1131	MOTIFS
					E1-E2 ATPases phosphorylation site: D438-T444	MOTIFS
3	7478871CD1	989	S23 S51 S65 S149 S261 S304 S309 S369 S795 S800 S936 S953 S966 S968 T158 T206 T336 T368 T388 T629 T656 T691 T864	N183 N555 N582 N606 N985	HCO3- transporter family domain: L222-I897, K108-V157	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3 (cont.)					Transmembrane Domains: P227-L247, G260-M280, D412-L440, Q448-I474, P501-F529, R531-L554, H628-T656, R665-K693, A724-A744, K756-A776, F825-M853, T895-G923 N-terminus is non-cytosolic	TMAP
					Anion exchangers family signature BL00219: G89-H120, Q224-L267, S269-R307, A308-K343, S382-A421, V422-D445, L475-F513, L515-I562, P631-D684, W721-L762, D763-R801, G806-L851, Y852-T895, I897-S936	BLIMPS_BLOCKS
					Anion exchangers family signatures: D372-Y424, A519-G571	PROFILESCAN
					Anion exchanger signature PR00165: F392-L414, Q417-G437, V450-G469,T473-S492, L504-S523, G536-L554,D632-L651, W719-M738	BLIMPS_PRINTS
					PROTEIN ANION EXCHANGE TRANSMEMBRANE BAND GLYCOPROTEIN LIPOPROTEIN PALMITATE BICARBONATE COTRANSPORTER PD001455: Q224-L846, S567-I897, L109-R189	BLAST_PRODOM
					BICARBONATE COTRANSPORTER SODIUM ELECTROGENIC NA+ PANCREAS COTRANSPORTER2 HCO3 TRANSPORTER F52B5.1 PD018437: Q898-N989	BLAST_PRODOM
					BAND 3 ANION TRANSPORT PROTEIN DM02294|P04920|602-1237: G620-E956, L318-P591, G187-G229	BLAST_DOMO

102

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7483601CD1	505	S41 S238 S465 T13 T53 T128 T234 T464 T503	N163 N166	Sulfate transporter family domain: L193-T503	HMMER_PFAM
					Transmembrane Domains: L93-I121, T128-I156, A179-G199, G212-V232, N258-F278, L286-G306, F336-K364, A417-I445, E468-A495 N-terminus is non-cytosolic	TMAP
					Sulfate transporters protein signature BL01130: S86-V139, S181-V232	BLIMPS_BLOCKS
					SULFATE TRANSPORTER TRANSPORT PROTEIN TRANSMEMBRANE GLYCOPROTEIN AFFINITY SULPHATE HIGH PERMEASE PD001121: I60-D155	BLAST_PRODOM
					PROTEIN TRANSPORT SULFATE TRANSPORTER TRANSMEMBRANE PERMEASE INTERGENIC REGION AFFINITY GLYCOPROTEIN PD001255: L257-R502	BLAST_PRODOM
					SULFATE TRANSPORTERS DM01229|P40879|5-462: R15-R463	BLAST_DOMO
5	7487851CD1	618	S127 S169 S259 S417 S458 S491 S590 S609 S616 T321 T522 T537	N167	Xanthine/uracil permeases family domain: G46-E481	HMMER_PFAM
					Transmembrane Domains: P44-C72, P198-L214, C224-G246, L267-P295, L319-Y343, L364-T383, S400-R419, L424-Y452, A454-A482, D494-E516 N-terminus is non-cytosolic	TMAP

103

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5 (cont.)					Xanthine/uracil permease signature BL01116: R362-G413, G415-F451	BLIMPS_BLOCKS
					YOLK SAC PERMEASELIKE YSPL1 FORM 1 YOLK SAC PERMEASELIKE YSPL1 FORM 4 YOLK SAC PERMEASELIKE YSPL1 FORM 3 YOLK SAC PERMEASELIKE YSPL1 FORM 2 PD019501: G437-Q617 PD137940: Q29-P83	BLAST_PRODOM
					XANTHINE/URACIL PERMEASES FAMILY DM01485|S33349|7-188: G363-L473	BLAST_DOMO
6	7472881CD1	377	S15 S16 S91 S324 S337 T310 T332 T336 T374	N4 N14 N157	Sodium Bile acid symporter family domain: T39-W220	HMMER_PFAM
					Signal Peptide: M41-A97	SPSCAN
					Transmembrane domains: G28-R56 A69-S89 V95-F115 T131-S153 T159-V182 K191-G218 W220-T248 L283-A30	TMAP
					PROTEIN TRANSMEMBRANE ACID COTRANSPORTING POLYPEPTIDE TRANSPORT SYMPORT SODIUM/BILE COTRANSPORTER NA+/BILE PD002890: M41-D223	BLAST_PRODOM
					ACID COTRANSPORTING POLYPEPTIDE SODIUM/BILE COTRANSPORTER NA+/BILE SODIUM/TAUROCHOLATE TRANSMEMBRANE TRANSPORT SYMPORT PD007533: W220-R313	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6 (cont.)					do SODIUM; ACID; BILE; TRANSPORTER; DM03972|I38655|8-318: L30-K321 DM03972|P09131|163-477: P12-S277 DM03972|P26435|1-314: A10-R313	BLAST_DOMO
7	7612560CD1	507	S22 S26 S41 S261 S341 S374 S384 T36	N181 N190 N232 N477	Transmembrane amino acid transporter protein domain: A78-G458	HMMER_PFAM
					Transmembrane domains: G74-M102, A143-F168, F208-L236, P266-E286, P296-L316, V342-I370, L381-P401, I407-E427, S437-A462 N-terminus is cytosolic	TMAP
					ACID AMINO PROTEIN TRANSPORTER PERMEASE TRANSMEMBRANE INTERGENIC REGION PUTATIVE PROLINE PD001875: K49-L356	BLAST_PRODOM
8	2880370CD1	438	S48 S80 S300 S407 T15 T38 T92	N56 N85 N99	Signal Peptide: M1-R20, M1-M21, M1-S23	HMMER
					Signal Cleavage: M1-A19	SPSCAN
					Sodium Bile acid symporter family: L148-D332	HMMER_PFAM
					Transmembrane domains: K4-R20, A135-F158, I178-A206, G218-M238, L244-S264, P270-V290, I305-G325, E335-A355, V368-P389, P400-R423 N-terminus is cytosolic	TMAP
					PROTEIN TRANSMEMBRANE ACID COTRANSPORTING POLYPEPTIDE TRANSPORT SYMPORT SODIUM/BILE COTRANSPORTER NA+/BILE PD002890: L150-D332	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8 (cont.)					P3 PROTEIN TRANSMEMBRANE TRANSPORT SYMPORT PD103884: G317-L416	BLAST_PRODOM
					do SODIUM; ACID; BILE; TRANSPORTER; DM03972|P09131|163-477: V121-L416 DM03972|I38655|8-318: I143-C424 DM03972|P26435|1-314: I143-R423	BLAST_DOMO
9	6267489CD1	350	S68 S121 S188 S233 S336 T29 T41 T136 T146 T288 Y84 Y194 Y241 Y294	N60 N87	ATP synthase (C/AC39) subunit: Y15-P348	HMMER_PFAM
					Transmembrane domain: R86-N114 N-terminus is non-cytosolic	TMAP
					SUBUNIT VATPASE AC39 VACUOLAR ATP SYNTHASE HYDROLASE HYDROGEN ION TRANSPORT PD008622: L78-G285	BLAST_PRODOM
					SUBUNIT VATPASE AC39 VACUOLAR ATP SYNTHASE HYDROLASE HYDROGEN ION TRANSPORT PD013947: L2-R77	BLAST_PRODOM
					do AC39; ATP; VACUOLAR; SYNTHASE DM03240|P54641|10-355: G4-I349 DM03240|P12953|1-272: E81-I349 DM03240|P53659|1-363: L2-G286; G201-I349 DM03240|P32366|32-344: N35-I349	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7484777CD1	1707	S54 S63 S80 S116 S122 S134 S150 S365 S388 S453 S519 S554 S681 S711 S771 S840 S841 S900 S1037 S1170 S1212 S1213 S1222 S1229 S1241 S1278 S1393 S1397 S1398 S1405 S1501 S1546 S1595 S1612 S1619 S1639 S1655 S1657 S1668 S1678 S1679 S1689 S1694 T42 T162 T300 T575 T612 T613 T1070 T1115 T1137 T1184 T1265 T1271 T1285 T1308 T1451 T1465 T1608 T1650 Y70 Y798 Y1010	N40 N111 N297 N386 N451 N573 N729 N732 N942 N1068 N1113 N1211 N1227 N1626	Transient receptor: Y1096-M1154, R970-E1035, P899-L960, D715-W761	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10 (cont.)					Transmembrane domain: W5-E27, G204-I228, D550-R578, F865-V893, L937-R959, V975-G995, M1005-A1025, W1087-T1115 N-terminus is non-cytosolic	TMAP
					Transient receptor potential family signature PR01097: A1094-T1115, F1116-F1129, V1143-M1156	BLIMPS_PRINTS
					PROTEIN MELASTATIN CHROMOSOME TRANSMEMBRANE C05C12.3 T01H8.5 I F54D1.5 IV PD018035: M154-L486	BLAST_PRODOM
					PROTEIN CHROMOSOME TRANSMEMBRANE MELASTATIN C05C12.3 T01H8.5 I F54D1.5 IV PD151509: I982-L1270	BLAST_PRODOM
					PROTEIN CHROMOSOME TRANSMEMBRANE MELASTATIN C05C12.3 T01H8.5 I F54D1.5 IV PD039592: E617-E813	BLAST_PRODOM
					PROTEIN MELASTATIN CHROMOSOME TRANSMEMBRANE T01H8.5 I C05C12.3 F54D1.5 IV PD022180: W481-R591	BLAST_PRODOM
					ANK MOTIF REPEAT DM03196|P34586|38-822: I972-C1162 DM03196|P19334|1-772:D962-I1157 DM03196|P48994|13-780:I978-Q1159	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	2493969CD1	771	S34 S156 S186 S379 S403 S435 S468 S488 S499 S677 S682 S703 S716 S744 T6 T54 T126 T273 T274 T449 T518 T543 T712	N163 N282 N676	Transmembrane domains: L49-G76 L77-Y105 V125-A153 S186-I211 G212-Y240 S252-T274 P286-Y314 G330-L350 F355-A375 I389-L417 T561-Y589 S594-P622 A629-K649 W655-W675 N-terminus is cytosolic	TMAP
					Amino acid permeases protein signature BL00218: V56-G84, V87-S118, Y263-L307, A344-T383	BLIMPS_BLOCKS
					AMINO ACID CATIONIC TRANSPORTER TRANSPORT TRANSMEMBRANE GLYCOPROTEIN TRANSPORTER1 PROTEIN HIGH AFFINITY PD000262: V614-L688	BLAST_PRODOM
					TRANSMEMBRANE TRANSPORT PROTEIN TRANSPORTER AMINO ACID PERMEASE AMINO ACID GLYCOPROTEIN MEMBRANE PD000214: L49-L421	BLAST_PRODOM
					do ANTIPORTER; ORNITHINE; PUTRESCINE; TRANSPORT; DM01125|P30825|23-373: T47-W241	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	3244593CD1	1329	S10 S20 S28 S81 S156 S208 S216 S230 S397 S407 S448 S473 S491 S517 S619 S631 S667 S725 S853 S868 S979 S1024 S1086 S1128 S1159 S1190 S1228 S1259 T152 T295 T301 T324 T373 T425 T452 T483 T575 T649 T684 T752 T805 T857 T875 T1046 T1055 T1091 T1180 T1268 Y714	N405 N438 N540 N602 N803 N951 N1226	ABC transporter transmembrane region: V123-I391, L766-V1044	HMMER_PFAM
					ABC transporter domain: G1117-G1300, G506-G677	HMMER_PFAM
					Transmembrane domains: F118-H146 V159-F179 A185-N205 E233-A253 G260-M280 A350-R370 S379-K399 T759-L786 H819-T846 F904-F932 N989-S1017 N-terminus is non-cytosolic	TMAP
					ABC transporters family signature: L585-D634, T1208-D1258	PROFILESCAN
					ATP-BINDING TRANSPORT TR PD00131: G876-D885, S1128-V1181, G1275-A1312	BLIMPS_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12 (cont.)					ATP-BINDING TRANSPORT TRANSMEMBRANE PROTEIN GLYCOPROTEIN MULTIDRUG SULFONYLUREA RECEPTOR RESISTANCE ASSOCIATED CONDUCTANCE PD003781: L543-L601	BLAST_PRODOM
					ABC TRANSPORTERS FAMILY DM00008|P33527|1293-1502: I1090-G1300, D490-G677	BLAST_DOMO
					ABC transporters family signature: L603-V617, F1227-L1241	MOTIFS
					ATP/GTP-binding site motif A (P-loop): G513-S520 G1124-S1131	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	4921451CD1	1353	S11 S53 S146 S183 S199 S347 S422 S500 S513 S532 S592 S638 S644 S841 S865 S876 S900 S1090 S1232 S1236 S1244 S1248 S1287 S1295 S1302 S1321 T8 T79 T113 T234 T306 T312 T391 T618 T639 T690 T744 T757 T807 T924 T1030 T1272 T1284 Y367 Y431 Y706	N637	Transmembrane domains: F130-L158 D394-S422 V448-L473 R996-A1024 F1055-R1083 D1093-V1113 I1117-I1137 S1163-I1191 N-terminus is non-cytosolic	TMAP
					E1-E2 ATPases phosphoryl BL00154: V508-F526, D748-L788, T943-A966	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: A494-P539	PROFILESCAN
					P-type cation-transporting atpase superfamily signature PR00119: F512-F526, S764-D774, I946-L965	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13 (cont.)					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATP-BINDING PROTEIN PROBABLE CALCIUM TRANSPORTING CALCIUM TRANSPORT PD004657: L981-V1034, G1028-I1180 PD006317: A270-D343, F200-P223 PD149930: C920-F979	BLAST_PRODOM
					PROBABLE CALCIUM TRANSPORTING ATPASE 8 EC 3.6.1.38 HYPOTHETICAL PROTEIN HYDROLASE CALCIUM TRANSPORT TRANSMEMBRANE PHOSPHORYLATION MAGNESIUM ATP-BINDING PD101227: G582-I768	BLAST_PRODOM
					do ATPASE; CALCIUM; TRANSPORTING; DM02405|P32660|318-1225: A270-E549, P580-L796, R906-G1031, F200-P223	BLAST_DOMO
					E1-E2 ATPases phosphorylation site: D514-T520	MOTIFS
					EF-hand calcium-binding domain: D1033-L1045	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	5547443CD1	921	S5 S46 S74 S215 S225 S277 S304 S475 S495 S502 S515 S538 S598 S656 S688 S747 S808 S829 S855 S881 T42 T57 T67 T127 T163 T329 T337 T364 T609 T614 T686 T710 T722 T781 T839 Y529 Y880	N223 N612	K+ channel tetramerisation domain: D8-H105, Q391-S488	HMMER_PFAM
					do CHANNEL; POTASSIUM; CDRK; SHAW; DM00490|P17971|32-138: N13-P92 (P-value = 8.5e-05)	BLAST_DOMO
15	56008413CD1	530	S6 S151 S268 S306 S476 T56 T57 T90 T199 T262 T338	N396 N523	Nucleoside transporter domain: L170-S507	HMMER_PFAM
					TRANSMEMBRANE DOMAINS: R66-Y94 G101-R129 T134-R162 T231-R256 V348-E375 H380-L408 H416-Y436 A447-P467 N-terminus is non-cytosolic	TMAP
					PROTEIN NUCLEOSIDE TRANSPORTER TRANSMEMBRANE NUCLEOLAR HNP36 DELAYED EARLY RESPONSE DER12 NUCLEAR PD005103: V182-Y503	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	6127911CD1	1617	S30 S50 S134 S249 S353 S491 S672 S761 S792 S809 S819 S915 S923 S954 S1035 S1127 S1193 S1269 S1295 S1329 S1488 T111 T206 T558 T572 T624 T643 T755 T772 T780 T852 T968 T1172 T1257 T1340 T1370 T1418 T1441 T1462 T1545 T1605 Y947	N71 N84 N91 N109 N130 N241 N436 N544 N576 N911 N940 N990 N1305	Signal Peptide: M26-L46	HMMER
					ABC transporter domains: G507-G689, G1313-G1489	HMMER_PFAM
					TRANSMEMBRANE DOMAINS: R25-N53 E221-K247 A262-I282 I292-V312 L322-L342 E356-N382 D392-I420 L848-Y876 H1006-G1034 Q1061-Y1081 V1095-M1115 F1132-V1160 C1200-M1226 N-terminus is non-cytosolic	TMAP
					ABC transporters family signature: V595-D646	PROFILESCAN
					ABC TRANSPORTERS FAMILY DM00008|P41233|839-1045: I478-N688, K1300-M1486	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16 (cont.)					ATP/GTP-binding site motif A (P-loop): G514-S521, G1320-S1327	MOTIFS
17	6427133CD1	1192	S4 S152 S216 S259 S268 S296 S366 S391 S408 S437 S440 S456 S483 S493 S545 S744 S833 S1114 S1115 S1124 S1125 S1144 S1157 S1168 T35 T267 T378 T403 T519 T540 T646 T900 T1063 T1095 T1120 T1178 T1189 Y22 Y28 Y607	N238 N538 N726 N1165	TRANSMEMBRANE DOMAINS: A58-L86 D270-W298 F327-H353 G862-F890 T900-G923 F950-Y978 A995-S1015 H1022-N1042 S1061-K1089	TMAP
					E1-E2 ATPases phosphorylation site signature BL00154: G133-L150, I386-F404, D650-M690, T810-S833	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: A372-L421	PROFILESCAN
					P-type cation-transporting atpase superfamily signature PR00119: F390-F404, A666-D676, I813-I832	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17 (cont.)					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATPBINDING PROTEIN PROBABLE CALCIUMTRANSPORTING CALCIUM TRANSPORT PD004657: S847-P1094 PD006317: Q123-H222 PD149930: C787-Y846	BLAST_PRODOM
					FIC1 PROTEIN PD180313: H1040-P1154	BLAST_PRODOM
					do ATPASE; CALCIUM; TRANSPORTING; DM02405|P39524|236-1049: L66-N696, A755-N911	BLAST_DOMO
					E1-E2 ATPases phosphorylation site: D392-T398	MOTIFS
18	7472932CD1	625	S86 S280 S339 S510 S554 T205 T387 T505 T516 T589 T594 T612	N144 N168 N174 N351	Sodium: neurotransmitter symporter family domain: R18-L588	HMMER_PFAM
					TRANSMEMBRANE DOMAINS: E17-R43 C48-L76 Y96-W124 S178-V198 T204-L224 P251-N279 V295-N323 P394-T414 E420-A440 C446-E466 A472-Y492 W513-R541 P561-T589 N-terminus is non-cytosolic	TMAP
					Sodium: neurotransmitter symporter family signature BL00610: Q26-E75, W90-C139, W181-G232, I247-T299, T389-V431, V485-P539, K558-P580	BLIMPS_BLOCKS
					Sodium: neurotransmitter symporter family signatures: D22-L76	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18 (cont.)					Sodium/neurotransmitter symporter signature PR00176: Q26-L47, A55-V74, G99-Y125, V208-I225, V290-V310, M393-L412, S474-M494, R514-L534	BLIMPS_PRINTS
					TRANSPORTER NEUROTRANSMITTER TRANSPORT TRANSMEMBRANE SYMPORT GLYCOPROTEIN SODIUM CHLORIDE-DEPENDENT SODIUM-DEPENDENT GABA PD000448: L363-R598, R18-D284	BLAST_PRODOM
					ORPHAN TRANSPORTER ISOFORM A12 A11 B11 A8 B9 A10 RENAL PD037829: K314-L368 PD150276: S137-Q180	BLAST_PRODOM
					TRANSMEMBRANE TRANSPORT PROTEIN TRANSPORTER AMINOACID PERMEASE AMINO ACID GLYCOPROTEIN MEMBRANE PD000214: L28-L311, S375-L534	BLAST_PRODOM
					SODIUM: NEUROTRANSMITTER SYMPORTER FAMILY DM00572|S50998|19-616:A11-R591	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	8463147CD1	1181	S16 S43 S52 S68 S97 S106 S153 S164 S196 S293 S347 S393 S424 S425 S531 S651 S674 S697 S709 S854 S907 S937 S973 S996 S1009 S1022 S1060 S1068 S1075 S1093 S1162 S1166 S1175 T81 T337 T377 T432 T503 T602 T701 T702 T977 T1013 T1046 T1137 T1171	N104 N137 N329 N591 N600 N619 N1044 N1169	TRANSMEMBRANE DOMAINS: L98-L120 W135-L163 I173-F201 K233-L259 L266-D286 V298-Y318 M808-S826 V867-T883 S918-Y944 N-terminus is cytosolic	TMAP
					CHANNEL POTASSIUM IONIC CALCIUM-ACTIVATED ALPHA CALCIUM SUBUNIT ACTIVATED PROTEIN LARGE PD003090: R323-F609, S877-P966, S656-G716, V771-V867, Q1123-I1148	BLAST_PRODOM
					do CHANNEL; POTASSIUM; MSLO; ACTIVATED; DM05442|A48206|351-1123: R323-F609, P927-P966, G777-V867, Q1123-I1148, G1110-E1160	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G1071-T1078	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7506408CD1	233	S71 S116 S219 T29 T171 Y77 Y124 Y177		ATP synthase (C/AC39) subunit: Y15-P231	HMMER_PFAM
					SUBUNIT VATPASE AC39 VACUOLAR ATP SYNTHASE HYDROLASE HYDROGEN ION TRANSPORT PD008622: G84-I232, G14-G168	BLAST_PRODOM
					ATP; VACUOLAR; SYNTHASE DM03240|P12953|1-272: F46-I232 DM03240|PS4641|10-355: D32-I232, G4-E43 DM03240|P53659|1-363: G14-I232 DM03240|P32366|32-344: V37-I232	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
21/6911460CB1/2232	1-512, 1-756, 5-607, 120-658, 144-421, 144-540, 144-598, 144-624, 144-646, 144-681, 144-694, 144-697, 144-727, 144-745, 144-769, 144-810, 147-764, 215-899, 320-522, 321-601, 321-681, 321-799, 321-817, 321-875, 321-884, 321-888, 321-899, 321-923, 322-969, 337-1058, 371-1112, 382-1044, 454-1062, 495-1011, 513-1350, 568-1324, 578-1130, 597-1205, 599-1249, 605-1124, 619-834, 620-1356, 641-1524, 674-1455, 678-1410, 689-1516, 701-1364, 724-1260, 731-1350, 731-1429, 731-1571, 732-1306, 738-1372, 748-1439, 750-1430, 761-1528, 765-1440, 772-1332, 785-1275, 806-1463, 819-1363, 843-1439, 848-1458, 875-1417, 916-1528, 918-1408, 928-1366, 928-1456, 931-1387, 945-1443, 948-1470, 953-1466, 955-1357, 956-1597, 957-1656, 962-1721, 967-1616, 969-1596, 1018-1630, 1031-1696, 1032-1730, 1034-1769, 1038-1724, 1060-1792, 1142-1787, 1163-1817, 1178-1848, 1179-1845, 1180-1765, 1182-1806, 1258-1586, 1258-1672, 1258-1679, 1258-1836, 1258-1840, 1258-1848, 1258-1851, 1258-1867, 1258-1881, 1258-1909, 1258-1938, 1258-1961, 1258-1967, 1320-1561, 1391-1679, 1504-1746, 1504-1788, 1504-1850, 1504-1852, 1504-1863, 1504-1872, 1504-1887, 1504-1888, 1504-1892, 1504-1893, 1504-1900,
	1504-1905, 1504-1912, 1504-1913, 1504-1920, 1504-1935, 1504-1947, 1504-1953, 1504-1960, 1504-1990, 1504-2014, 1504-2021, 1504-2031, 1504-2051, 1504-2070, 1504-2089, 1504-2098, 1504-2127, 1504-2202, 1504-2217, 1506-2175, 1573-2219, 1576-1698, 1579-2222, 1579-2232, 1581-2231, 1583-2212, 1591-2132, 1623-1914, 1649-1744, 1652-1955
22/55138203CB1/4135	1-735, 3-735, 5-729, 5-735, 21-735, 37-735, 87-516, 310-610, 310-758, 310-831, 310-849, 518-1026, 529-735, 533-1026, 580-735, 685-735, 687-735, 745-1412, 754-1188, 1159-1631, 1159-1640, 1561-1938, 1700-1868, 1875-2532, 2221-2367, 2251-2460, 2251-2540, 2368-2835, 2488-3083, 2512-3140, 2544-3085, 2580-3188, 2724-3031, 2724-3131, 3112-3707, 3113-3165, 3113-3429, 3484-4079, 3556-3737, 3571-3776, 3571-4135, 3612-4095, 3614-3733, 3648-4098, 3649-4079, 3650-3717
23/7478871CB1/2970	1-302, 1-462, 303-462, 406-576, 406-706, 648-2970, 649-809, 649-897, 810-897, 810-1052, 898-1052, 898-1169, 1053-1169, 1170-1344, 1170-1478, 1255-2103, 1345-1478, 1345-1742, 1479-1742, 1479-1800, 1563-1800, 1623-1800, 1764-1800, 1801-1989, 1801-2103, 1990-2103, 1990-2265, 2104-2265, 2104-2444, 2107-2343, 2266-2444, 2266-2517, 2445-2517, 2518-2586, 2518-2760, 2587-2760, 2587-2916, 2760-2970, 2761-2916, 2917-2970
24/7483601CB1/1835	1-152, 1-292, 1-890, 34-669, 34-673, 36-673, 61-673, 153-292, 153-403, 293-403, 293-570, 404-570, 569-735, 569-890, 591-888, 736-890, 820-1233, 820-1543, 820-1578, 820-1617, 827-1233, 1309-1643, 1309-1835

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
25/7487851CB1/2220	1-232, 1-367, 1-427, 1-517, 1-578, 1-625, 1-703, 1-854, 4-427, 4-558, 5-297, 5-397, 7-686, 79-631, 185-717, 233-497, 241-862, 265-808, 271-427, 271-571, 271-670, 271-680, 271-772, 271-775, 271-792, 271-796, 271-812, 271-836, 271-886, 271-895, 271-955, 273-961, 273-1104, 277-807, 292-947, 293-947, 323-641, 342-935, 353-947, 367-427, 382-427, 386-903, 395-427, 397-427, 400-588, 451-657, 451-708, 452-485, 452-489, 452-670, 452-886, 452-947, 486-947, 489-1046, 493-662, 507-695, 540-1351, 564-1150, 577-1232, 577-1359, 579-1233, 581-947, 586-947, 590-947, 612-1225, 621-708, 647-947, 701-1369, 708-833, 709-947, 734-835, 734-898, 741-833, 746-947, 758-1355, 758-1451, 764-947, 764-1424, 765-1374, 771-1285, 776-947, 777-1530, 790-1014, 799-1312, 810-947, 817-1398, 828-1375, 831-948, 841-1505, 845-1253, 859-1423, 861-1026, 876-1364, 877-1502, 880-1382, 888-1602, 891-1715, 907-1418, 926-1516, 961-1113, 971-1373, 973-1277, 980-1418, 991-1161, 1010-1514, 1016-1568, 1032-1563, 1033-1309, 1055-1863, 1057-1686, 1059-1200, 1062-1584, 1069-1871, 1070-1241, 1078-1675, 1081-1565,
	1081-1578, 1104-1738, 1112-1595, 1133-1895, 1154-1682, 1157-1762, 1172-1778, 1176-1370, 1184-1918, 1192-1463, 1197-1301, 1197-1306, 1201-1306, 1202-1709, 1238-1629, 1250-2078, 1251-2011, 1257-1306, 1259-1306, 1263-1828, 1266-1306, 1266-1897, 1268-1752, 1283-1306, 1289-1789, 1291-1784, 1301-1810, 1304-1660, 1306-1676, 1319-1716, 1319-1721, 1351-1660, 1358-1382, 1358-1497, 1358-1527, 1358-1558, 1358-1636, 1358-1660, 1358-1667, 1358-1693, 1358-1717, 1358-1738, 1358-1760, 1358-1771, 1358-1843, 1358-1899, 1358-1952, 1358-1971, 1358-1974, 1358-1995, 1358-1998, 1358-2044, 1361-1920, 1363-2057, 1365-1882, 1365-2051, 1377-1963, 1380-1975, 1421-2090, 1427-2012, 1454-2037, 1473-1613, 1482-1952, 1491-1630, 1501-2125, 1519-2153, 1519-2178, 1530-2216, 1532-2194, 1548-1955, 1554-1818, 1557-1795, 1557-2123, 1566-2219, 1567-2160, 1568-1863, 1570-2053, 1570-2220, 1577-2078, 1581-2220, 1597-2220, 1601-2200, 1606-2192, 1608-2220, 1623-2220, 1631-1946, 1647-2176, 1651-1931, 1651-2107, 1651-2176, 1651-2219, 1651-2220, 1654-1945, 1671-2220, 1672-1841, 1674-2220, 1690-2220, 1702-2106, 1728-1941, 1754-2220, 1794-2220, 1796-2220, 1835-2220, 1845-2090,
	1845-2220, 1867-2220, 1871-2220, 1874-2220, 1878-2220, 1898-2220, 1900-2220, 1954-2220, 2019-2220, 2037-2220, 2052-2220, 2078-2220, 2094-2220
26/7472881CB1/1517	1-236, 47-219, 134-622, 243-1070, 245-821, 245-841, 245-888, 245-901, 245-935, 245-951, 245-988, 249-622, 250-1073, 292-1071, 292-1073, 309-1073, 347-1073, 385-621, 385-622, 386-621, 418-621, 425-1073, 625-1383, 847-1515, 847-1517, 877-1517, 893-1073

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
27/7612560CB1/2142	1-258, 1-450, 1-502, 1-548, 1-575, 1-595, 1-599, 1-670, 1-679, 1-749, 13-814, 53-278, 53-493, 53-495, 53-502, 53-517, 53-546, 53-552, 53-563, 53-601, 53-604, 53-609, 53-614, 53-618, 56-597, 56-611, 56-613, 56-614, 195-983, 292-979, 301-832, 355-803, 452-1077, 501-1142, 533-845, 536-985, 552-1013, 615-1269, 615-1613, 624-880, 631-1030, 641-1231, 686-1269, 792-1269, 811-1269, 820-1269, 852-1266, 865-1269, 909-1269, 925-1269, 926-1269, 933-1321, 933-1326, 1026-1269, 1070-1269, 1081-1269, 1199-1537, 1538-1832, 1553-1806, 1553-2106, 1553-2131, 1553-2142
28/2880370CB1/1661	1-526, 1-528, 1-529, 345-1661, 465-673, 465-838, 465-854, 569-833, 1031-1307, 1032-1308, 1032-1590
29/6267489CB1/1501	1-280, 82-362, 100-369, 100-379, 103-742, 103-795, 112-571, 113-399, 124-653, 145-267, 182-441, 182-904, 514-588, 593-1215, 638-1313, 640-1140, 640-1149, 640-1152, 686-1314, 735-1225, 739-1287, 739-1289, 772-1358, 804-1501, 836-1498, 841-1294, 933-1501
30/7484777CB1/5526	1-658, 100-713, 100-738, 100-904, 100-931, 250-944, 414-556, 414-593, 414-600, 414-611, 414-616, 414-625, 414-703, 414-707, 414-724, 414-750, 414-856, 414-884, 414-886, 414-887, 414-903, 414-904, 414-911, 414-912, 414-919, 414-928, 414-929, 414-935, 414-939, 414-953, 414-961, 414-972, 414-974, 414-1008, 414-1022, 414-1032, 414-1043, 414-1048, 414-1064, 414-1065, 414-1077, 414-1084, 414-1108, 414-1118, 414-1154, 414-1179, 414-1180, 416-1014, 419-988, 431-563, 431-565, 432-1145, 454-886, 459-1154, 469-1095, 486-1018, 486-1033, 492-698, 502-1072, 522-966, 572-1219, 622-1337, 644-1123, 644-1211, 659-1329, 666-1155, 676-1331, 676-1332, 686-1418, 691-1332, 694-1332, 694-1333, 694-1359, 701-1155, 704-1333, 705-1333, 714-1333, 719-1398, 723-1333, 727-1478, 729-1333, 730-1285, 731-1155, 736-1426, 746-1333, 746-1419, 752-1418, 773-1419, 775-1333, 779-1333, 780-1419, 782-1333, 787-1333, 797-1333, 798-1333, 800-1480, 819-1384, 822-1515, 839-1332, 844-1333, 845-1333, 849-1039, 850-1613, 887-1333, 890-1333, 892-1333, 906-1333, 908-1138, 910-1384, 912-1333, 919-1333, 920-1154, 926-1384, 953-1516, 975-1592, 983-1384, 997-1399, 997-1419, 1007-1683, 1038-1525, 1038-1685,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
30 (cont.)	1038-1696, 1038-1699, 1038-1773, 1040-1699, 1186-1917, 1220-1898, 1222-1907, 1291-1979, 1374-1991, 1635-2139, 1635-2151, 1635-2198, 1635-2454, 1639-2311, 2018-2692, 2018-2725, 2081-2852, 2138-2817, 2169-2725, 2263-2929, 2283-2940, 2293-2991, 2302-3152, 2312-2786, 2338-2973, 2340-2887, 2351-2896, 2352-3152, 2365-3152, 2365-3170, 2382-2886, 2457-2996, 2568-3415, 2700-3483, 2705-3313, 2722-3373, 2746-3423, 2765-3236, 2770-3423, 2822-3530, 2823-3645, 2845-3703, 2854-3533, 2860-3423, 2868-3423, 2876-3423, 2880-3423, 2917-3347, 2946-3423, 2975-3218, 2975-3261, 2975-3359, 2975-3361, 2976-3352, 2976-3389, 2976-3393, 2976-3419, 2976-3506, 2976-3550, 2986-3478, 3010-3247, 3046-3414, 3142-3378, 3142-3600, 3143-3668, 3147-3859, 3236-3721, 3327-4170, 3696-4179, 3773-4176, 3773-4196, 3773-4242, 3773-4253, 3773-4271, 3773-4274, 3773-4276, 3773-4284, 3773-4290, 3773-4302, 3773-4303, 3773-4329, 3773-4337, 3773-4339, 3773-4340, 3773-4350, 3773-4354, 3773-4377, 3773-4427, 3773-4467, 3773-4480, 3773-4487, 3773-4505, 3773-4521, 3773-4567, 3773-4572, 3775-4299, 3775-4308, 3775-4478, 3786-4670, 3804-4253, 3819-4444, 3943-4822, 3964-4798, 3988-4445, 3989-
	4599, 4001-4179, 4204-4615, 4242-4500, 4251-5000, 4266-5000, 4309-5000, 4310-4696, 4329-5000, 4474-5000, 4797-5000, 4813-5000, 4870-5142, 4870-5335, 4870-5381, 4870-5388, 4870-5406, 4870-5432, 4870-5440, 4870-5441, 4870-5449, 4870-5462, 4870-5468, 4870-5515, 4870-5516, 4870-5526, 4872-5469, 4873-5245, 4946-5467, 4956-5403
31/2493969CB1/2739	1-701, 1-705, 43-383, 197-760, 293-2536, 980-1126, 1174-1443, 1218-1860, 1256-1863, 1339-1863, 1563-1880, 2001-2716, 2097-2715, 2156-2703, 2213-2311, 2223-2739, 2299-2739
32/3244593CB1/4321	1-1712, 32-979, 32-1712, 980-2810, 1089-1645, 1089-1661, 1089-1676, 1089-1700, 1089-1710, 1089-1711, 1711-3990, 1988-2016, 2282-2628, 2282-2631, 2282-2845, 2282-3545, 2285-2629, 2300-2845, 3103-3344, 3103-3395, 3103-3528, 3103-3540, 3103-3545, 3103-3573, 3103-3603, 3103-3613, 3103-3616, 3103-3620, 3103-3629, 3103-3660, 3103-3687, 3103-3708, 3103-3730, 3103-3754, 3103-3772, 3103-3778, 3103-3805, 3103-3809, 3103-3836, 3103-3856, 3103-3881, 3106-3789, 3115-3369, 3115-3586, 3119-3670, 3132-3545, 3132-3573, 3143-3417, 3143-3545, 3177-3545, 3235-3545, 3262-4127, 3275-3545, 3315-3545, 3318-3545, 3351-3545, 3355-3944, 3360-3545, 3366-3545, 3380-3545, 3384-3545, 3390-3771, 3397-3926, 3415-3545, 3438-3545, 3439-3484, 3444-3545, 3445-3545, 3477-3545, 3546-3804, 3546-3839, 3546-3843, 3546-3859, 3546-3866, 3546-3868, 3546-3874, 3546-3878, 3546-3884, 3546-3893, 3546-3904, 3546-3907, 3546-3927, 3546-3934, 3546-3937, 3546-3953, 3546-3954, 3546-3961, 3546-3966, 3546-3977, 3546-3989, 3546-3994, 3546-4050, 3546-4057, 3546-4065, 3546-4075, 3546-4103,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
32 (cont.)	3546-4136, 3546-4142, 3546-4214, 3552-4084, 3554-4157, 3554-4181, 3554-4229, 3555-4218, 3559-4075, 3564-4063, 3573-4079, 3606-4206, 3614-4188, 3633-4321, 3641-4321, 3661-4287, 3664-4321, 3666-4292, 3668-4279, 3669-4022, 3681-4306, 3683-4223, 3688-4291, 3708-4227, 3713-4243, 3716-4314, 3740-4185, 3753-4319, 3769-4287, 3781-4127, 3796-4321
33/4921451CB1/4519	1-246, 1-373, 158-323, 158-373, 258-672, 383-409, 383-466, 559-677, 559-751, 559-753, 559-986, 559-1073, 894-1524, 898-1382, 973-1484, 973-1555, 1046-4299, 1116-1556, 1181-1839, 1255-1529, 1308-1839, 1309-1838, 1343-1821, 1434-1814, 1440-1814, 1464-1834, 1488-1839, 1571-1839, 1709-1766, 1847-1987, 3331-3745, 3950-4069, 3950-4119, 3950-4160, 3950-4216, 3956-4297, 4067-4516, 4076-4519, 4166-4519, 4204-4492, 4242-4518, 4253-4516
34/5547443CB1/2922	1-297, 13-297, 96-364, 96-697, 126-297, 298-2778, 649-889, 749-986, 1071-1185, 1071-1744, 1491-1751, 1593-2023, 1820-2297, 1820-2309, 1820-2349, 1820-2352, 1981-2904, 2067-2319, 2087-2722, 2128-2840, 2173-2605, 2211-2843, 2236-2904, 2238-2863, 2259-2521, 2259-2873, 2259-2915, 2271-2838, 2492-2919, 2563-2746, 2587-2922
35/56008413CB1/2763	1-470, 1-499, 1-533, 1-569, 10-501, 26-330, 43-652, 43-678, 43-679, 43-756, 43-769, 43-779, 44-779, 51-779, 68-779, 322-779, 323-779, 418-779, 423-600, 428-600, 457-1236, 544-600, 587-779, 653-1152, 653-1161, 707-1313, 922-1618, 1014-1288, 1085-1444, 1094-1349, 1094-1705, 1100-1692, 1125-1439, 1278-1836, 1311-1611, 1311-1735, 1410-1483, 1459-2056, 1469-1657, 1471-1953, 1479-1758, 1502-2055, 1516-2176, 1518-1988, 1533-1657, 1557-2261, 1562-1689, 1614-2220, 1632-1924, 1675-1772, 1689-2242, 1689-2329, 1713-2332, 1722-1878, 1723-2264, 1729-1858, 1739-2273, 1743-2413, 1748-2276, 1757-2381, 1790-2381, 1794-2224, 1799-2273, 1807-2107, 1825-2381, 1853-2463, 1857-2434, 1868-2346, 1879-2385, 1882-1986, 1886-2134, 1886-2353, 1886-2376, 1886-2380, 1886-2381, 1886-2389, 1886-2393, 1886-2396, 1886-2404, 1886-2407, 1886-2414, 1886-2448, 1886-2455, 1886-2456, 1886-2461, 1886-2464, 1886-2525, 1886-2538, 1886-2543, 1886-2555, 1886-2582, 1886-2602, 1886-2666, 1888-2448, 1888-2526, 1889-2545, 1893-2207, 1893-2477, 1897-2135, 1897-2528, 1900-2109, 1902-2088,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
35 (cont.)	1919-2477, 1923-2585, 1928-2598, 1953-2325, 1954-2040, 1955-2162, 2019-2254, 2024-2526, 2038-2434, 2038-2443, 2038-2459, 2038-2496, 2038-2585, 2046-2763, 2051-2556, 2086-2728, 2133-2604, 2144-2622, 2151-2642, 2151-2651, 2159-2721, 2171-2760, 2172-2736, 2189-2438, 2194-2464, 2194-2642, 2194-2662, 2194-2686, 2194-2729, 2194-2747, 2194-2758, 2194-2760, 2194-2761, 2194-2762, 2207-2699, 2219-2701, 2219-2763, 2229-2667, 2237-2763, 2243-2714, 2245-2763, 2256-2507, 2260-2545, 2268-2546, 2280-2540, 2286-2763, 2291-2761, 2308-2738, 2328-2762, 2340-2605, 2346-2746, 2352-2745, 2354-2486, 2357-2746, 2367-2707, 2383-2592, 2392-2744, 2396-2763, 2400-2746, 2401-2594, 2404-2529, 2407-2746, 2433-2746
36/6127911CB1/5211	1-404, 1-442, 1-483, 1-510, 1-554, 1-562, 1-566, 1-581, 1-582, 1-597, 1-602, 1-604, 1-621, 1-627, 1-633, 1-634, 1-639, 1-640, 2-423, 7-317, 22-640, 26-640, 40-640, 44-503, 44-559, 53-582, 54-323, 88-640, 104-640, 138-640, 177-640, 248-640, 277-971, 466-640, 483-744, 549-640, 581-640, 641-696, 745-815, 745-972, 780-1353, 780-1363, 780-1380, 868-1300, 1035-1804, 1099-1637, 1114-2221, 1115-1300, 1130-1839, 1130-1880, 1242-1712, 1257-1865, 1273-1915, 1319-1972, 1356-1964, 1375-2003, 1391-1990, 1412-2119, 1453-1982, 1453-1983, 1453-2035, 1453-2092, 1453-2101, 1453-2130, 1467-2185, 1468-1863, 1479-1967, 1484-2118, 1501-2242, 1501-2391, 1589-1982, 1589-2159, 1589-2186, 1591-2160, 1613-2051, 1618-2037, 1618-2109, 1618-2118, 1618-2119, 1709-2485, 1710-2119, 1739-2236, 1745-2375, 1745-2406, 1794-2486, 1796-2485, 1805-2485, 1806-2485, 1840-2486, 1901-2702, 2263-2753, 2264-2553, 2338-2954, 2338-2962, 2417-2909, 2417-2987, 2417-2993, 2417-2998, 2417-3003, 2417-3029, 2417-3036, 2417-3050, 2453-2591, 2453-2753, 2453-2868, 2453-3050, 2459-2882, 2513-2920, 2531-2976,
	2547-2784, 2592-3110, 2602-2753, 2612-3212, 2618-3086, 2706-3502, 2754-3372, 2767-3266, 2767-3294, 2767-3326, 2767-3430, 2768-3353, 2771-3284, 2773-3386, 2774-3386, 2821-3307, 2821-3369, 2821-3386, 2822-3085, 2822-3386, 2824-3501, 2836-3495, 2844-3490, 2855-3503, 2859-3503, 2864-3503, 2870-3503, 2872-3503, 2874-3503, 2875-3331, 2875-3465, 2883-3070, 2883-3221, 2883-3252, 2883-3277, 2883-3319, 2883-3348, 2883-3403, 2883-3415, 2883-3424, 2883-3488, 2883-3503, 2886-3503, 2888-3503, 2892-3503, 2893-3503, 2894-3503, 2900-3503, 2906-3503, 2924-3457, 2924-3502, 2924-3503, 2926-3503, 2931-3503, 2948-3503, 2971-3503, 2974-3467, 2974-3475, 2974-3502, 2974-3503, 2979-3503, 2983-3476, 2983-3503, 2986-3503, 3000-3503, 3001-3503, 3025-3503, 3062-3503, 3096-3503, 3214-3503, 3260-3502, 3260-3503, 3271-3493, 3271-3686, 3271-3932, 3341-3616, 3341-3787, 3341-3821, 3341-3901, 3341-3943, 3341-3954, 3341-3968, 3341-4001, 3341-4003, 3341-4009, 3410-3503, 3417-4146, 3539-4192, 3550-4269, 3596-4177, 3680-4297, 3683-4298, 3692-4186, 3695-4287, 3734-4536, 3765-4431, 3794-4495, 3816-4498, 3826-4329, 3838-4164, 3844-4094, 3847-4496, 3853-4511, 3855-4369, 3857-

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36 (cont.)	4430, 3860-4372, 3873-4548, 3876-4483, 3878-4427, 3896-4133, 3900-4238, 3900-4466, 3904-4455, 3905-4461, 3913-4467, 3914-4544, 4029-4763, 4104-4803, 4108-4793, 4109-4740, 4117-4794, 4125-4794, 4133-4851, 4139-4792, 4153-4851, 4160-4775, 4166-4834, 4169-4819, 4172-4812, 4179-4798, 4194-4867, 4210-4876, 4211-4805, 4216-4837, 4216-4909, 4219-4726, 4220-4807, 4220-4929, 4243-4749, 4245-4742, 4257-4806, 4258-4992, 4287-4900, 4287-5003, 4289-4926, 4292-4848, 4294-4542, 4299-4963, 4310-5014, 4312-4793, 4314-4879, 4319-4851, 4332-4883, 4355-4985, 4358-4978, 4361-4879, 4363-4976, 4365-4820, 4366-4979, 4370-4989, 4371-5125, 4380-5000, 4386-4985, 4393-4954, 4404-4797, 4405-5050, 4422-4873, 4427-5072, 4429-5090, 4430-5036, 4432-4975, 4434-4982, 4436-5082, 4450-5062, 4453-4977, 4457-4968, 4459-4878, 4459-5037, 4468-5071, 4482-5049, 4486-5175, 4500-5015, 4504-5204, 4512-5133, 4520-5094, 4522-5079, 4522-5087, 4523-4854, 4530-5211, 4540-5184, 4543-4768, 4544-5135, 4550-5109, 4554-4886, 4568-5043, 4579-4849, 4579-4978, 4579-5097, 4581-5131, 4582-4866, 4617-5087, 4632-4942, 4632-5104, 4632-5211, 4651-4847, 4651-4864, 4653-5138, 4658-4936, 4667-5092,
	4668-5211, 4799-5211
37/6427133CB1/5701	1-659, 1-716, 1-725, 1-741, 1-782, 22-518, 209-820, 275-532, 522-595, 535-595, 672-899, 688-1061, 688-1184, 688-1240, 688-1256, 747-1256, 908-996, 996-1193, 1002-1236, 1002-1612, 1039-1256, 1112-1256, 1153-1256, 1189-1612, 1196-1263, 1197-1446, 1447-1613, 1447-1917, 1447-2036, 1910-2173, 1910-2594, 2193-2300, 2301-2856, 2631-2744, 2669-2952, 2670-2856, 2670-2953, 2757-3430, 2802-2856, 2857-2959, 2860-3419, 2938-3210, 2946-3493, 3097-3704, 3097-3763, 3349-3996, 3520-3793, 3636-3884, 3707-3988, 3707-4166, 3867-4396, 3878-4150, 4026-4615, 4071-4615, 4086-4615, 4087-4554, 4139-4665, 4214-4496, 4290-4890, 4405-4766, 4476-4928, 4487-4772, 4499-4781, 4499-4788, 4499-5052, 4518-4774, 4555-4936, 4635-4905, 4635-4910, 4642-4803, 4693-4922, 4711-4992, 4778-5276, 4800-5384, 4855-5129, 4930-5158, 4930-5393, 4939-5436, 4949-5222, 5000-5693, 5000-5701, 5083-5693, 5102-5679, 5112-5701, 5118-5693, 5122-5674, 5122-5684, 5150-5452, 5163-5664, 5229-5497, 5232-5701, 5233-5693, 5292-5680, 5369-5619, 5503-5693, 5601-5694, 5624-5697
38/7472932CB1/1990	1-935, 1-1122, 954-1122, 965-1788, 967-1787, 1361-1485, 1505-1983, 1541-1987, 1549-1990, 1570-1989, 1586-1985, 1681-1988, 1788-1839, 1824-1875
39/8463147CB1/3760	1-204, 159-209, 170-243, 170-752, 290-389, 499-657, 674-1434, 675-1434, 767-1295, 769-1008, 769-1302, 773-1434, 800-1434, 1234-1764, 1234-1766, 1234-1772, 1303-1921, 1544-1725, 1544-2003, 1544-2058, 1922-2139, 2085-2139, 2090-2139, 2090-2414, 2242-2733, 2492-3052, 2656-3052, 2694-2971, 2694-3349, 2759-3349, 3049-3591, 3055-3349, 3233-3760, 3256-3658

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
40/7506408CB1/1150	1-280, 1-468, 1-560, 1-630, 1-1150, 101-200, 105-200, 110-200, 155-739, 178-520, 240-798, 256-789, 263-801, 266-962, 287-1072, 290-1072, 384-874, 388-936, 388-938, 422-1007, 490-943, 490-1147, 502-1072, 583-1150, 668-943

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
21	6911460CB1	BRAXTDR15
22	55138203CB1	THYMNOR02
23	7478871CB1	KIDNNOT32
25	7487851CB1	LUNGNOT37
26	7472881CB1	LIVRTUE01
27	7612560CB1	KIDCTME01
28	2880370CB1	ISLTNOT01
29	6267489CB1	KIDETXS02
30	7484777CB1	BRADDIR01
31	2493969CB1	BRAINOY02
32	3244593CB1	BRAENOT02
33	4921451CB1	PANCTUT01
34	5547443CB1	TESTNOT11
35	56008413CB1	LIVRTUE01
36	6127911CB1	BRSTNOT01
37	6427133CB1	TLYMNOT08
39	8463147CB1	BRAIFET02
40	7506408CB1	BONSTUT01

Table 6

Library	Vector	Library Description
BONSTUT01	pINCY	Library was constructed using RNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female during an exploratory laparotomy with soft tissue excision. Pathology indicated giant cell tumor of the sacrum. Patient history included a soft tissue malignant neoplasm. Family history included prostate cancer.
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAENOT02	pINCY	Library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure.
BRAFET02	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAINOY02	pINCY	This large size-fractionated and normalized library was constructed using pooled cDNA generated using mRNA isolated from midbrain, inferior temporal cortex, medulla, and posterior parietal cortex tissues removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. Scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver. 0.28 million independent clones from this size-selected library were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAXTDR15	PCDNA2.1	This random primed library was constructed using RNA isolated from superior parietal neocortex tissue removed from a 55 year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.

Table 6

Library	Vector	Library Description
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
KIDCTME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology indicated the margins of resection were free of involvement. Pathology for the matched tumor tissue indicated grade 3 renal cell carcinoma, clear cell type, forming a variegated multicystic mass situated within the mid-portion of the kidney. The tumor invaded deeply into but not through the renal capsule.
KIDETXS02	pINCY	This subtracted, transformed embryonal cell line library was constructed using 9 million clones from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue and was subjected to two rounds of subtraction hybridization with 1.9 million clones from an untreated transformed embryonal cell line (293-EBNA) derived from a kidney epithelial tissue library. The starting library for subtraction was constructed using RNA isolated from the treated, transformed embryonal cell line (293-EBNA). The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated 293-EBNA cells from the same cell line. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research (1996) 6:791.
KIDNNOT32	pINCY	Library was constructed using RNA isolated from kidney tissue removed from a 49-year-old Caucasian male who died from an intracranial hemorrhage and cerebrovascular accident. Patient history included tobacco abuse.
LIVRTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father.
LUNGNOT37	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 15-year-old Caucasian female who died from a closed head injury. Serology was positive for cytomegalovirus.

Table 6

Library	Vector	Library Description
PANCTUT01	pINCY	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
TESTNOT11	pINCY	Library was constructed using RNA isolated from testicular tissue removed from a 16-year-old Caucasian male who died from hanging. Patient history included drug use (tobacco, marijuana, and cocaine use), and medications included Lithium, Ritalin, and Paxil.
THYMNOR02	pINCY	The library was constructed using RNA isolated from thymus tissue removed from a 2-year-old Caucasian female during a thymectomy and patch closure of left atrioventricular fistula. Pathology indicated there was no gross abnormality of the thymus. The patient presented with congenital heart abnormalities. Patient history included double inlet left ventricle and a rudimentary right ventricle, pulmonary hypertension, cyanosis, subaortic stenosis, seizures, and a fracture of the skull base. Family history included reflux neuropathy.
TLYMNOT08	pINCY	The library was constructed using RNA isolated from anergicallogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male.The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM and SMART.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM or SMART hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seatle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
   a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,
   b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,
   c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and
   d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:
   a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide

encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

19. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and

a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides

of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

<110> INCYTE GENOMICS, INC.
LEE, Ernestine A.
DING, Li
BAUGHN, Mariah R.
TRIBOULEY, Catherine M.
BRUNS, Christopher M.
ELLIOTT, Vicki S.
WALIA, Narinder K.
FORSYTHE, Ian
RAUMANN, Brigitte E.
BURFORD, Neil
LAL, Preeti G.
THORNTON, Michael
GANDHI, Ameena R.
ARIVZU, Chandra
YAO, Monique G.
YUE, Henry
XU, Yuming
HAFALIA, April J.A.
ISON, Craig H.
CHEN, Huei-Mei

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0356 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/267,892; 60/271,168; 60/272,890; 60/276,860; 60/278,255; 60/280,538; 60/351,359
<151> 2001-02-09; 2001-02-23; 2001-03-02; 2001-03-16; 2001-03-23; 2001-03-30; 2002-01-25

<160> 40
<170> PERL Program

<210> 1
<211> 617
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6911460CD1

<400> 1
```
Met Val Pro Val Glu Asn Thr Glu Gly Pro Ser Leu Leu Asn Gln
  1               5                  10                  15
Lys Gly Thr Ala Val Glu Thr Glu Gly Ser Gly Ser Arg His Pro
             20                  25                  30
Pro Trp Ala Arg Gly Cys Gly Met Phe Thr Phe Leu Ser Ser Val
             35                  40                  45
Thr Ala Ala Val Ser Gly Leu Leu Val Gly Tyr Glu Leu Gly Ile
             50                  55                  60
Ile Ser Gly Ala Leu Leu Gln Ile Lys Thr Leu Leu Ala Leu Ser
```



```
                 65                  70                  75
Cys His Glu Gln Glu Met Val Val Ser Ser Leu Val Ile Gly Ala
             80                  85                  90
Leu Leu Ala Ser Leu Thr Gly Gly Val Leu Ile Asp Arg Tyr Gly
             95                 100                 105
Arg Arg Thr Ala Ile Ile Leu Ser Ser Cys Leu Leu Gly Leu Gly
            110                 115                 120
Ser Leu Val Leu Ile Leu Ser Leu Ser Tyr Thr Val Leu Ile Val
            125                 130                 135
Gly Arg Ile Ala Ile Gly Val Ser Ile Ser Leu Ser Ser Ile Ala
            140                 145                 150
Thr Cys Val Tyr Ile Ala Glu Ile Ala Pro Gln His Arg Arg Gly
            155                 160                 165
Leu Leu Val Ser Leu Asn Glu Leu Met Ile Val Ile Gly Ile Leu
            170                 175                 180
Ser Ala Tyr Ile Ser Asn Tyr Ala Phe Ala Asn Val Phe His Gly
            185                 190                 195
Trp Lys Tyr Met Phe Gly Leu Val Ile Pro Leu Gly Val Leu Gln
            200                 205                 210
Ala Ile Ala Met Tyr Phe Leu Pro Pro Ser Pro Arg Phe Leu Val
            215                 220                 225
Met Lys Gly Gln Glu Gly Ala Ala Ser Lys Val Leu Gly Arg Leu
            230                 235                 240
Arg Ala Leu Ser Asp Thr Thr Glu Glu Leu Thr Val Ile Lys Ser
            245                 250                 255
Ser Leu Lys Asp Glu Tyr Gln Tyr Ser Phe Trp Asp Leu Phe Arg
            260                 265                 270
Ser Lys Asp Asn Met Arg Thr Arg Ile Met Ile Gly Leu Thr Leu
            275                 280                 285
Val Phe Phe Val Gln Ile Thr Gly Gln Pro Asn Ile Leu Phe Tyr
            290                 295                 300
Ala Ser Thr Val Leu Lys Ser Val Gly Phe Gln Ser Asn Glu Ala
            305                 310                 315
Ala Ser Leu Ala Ser Thr Gly Val Gly Val Val Lys Val Ile Ser
            320                 325                 330
Thr Ile Pro Ala Thr Leu Leu Val Asp His Val Gly Ser Lys Thr
            335                 340                 345
Phe Leu Cys Ile Gly Ser Ser Val Met Ala Ala Ser Leu Val Thr
            350                 355                 360
Met Gly Ile Val Asn Leu Asn Ile His Met Asn Phe Thr His Ile
            365                 370                 375
Cys Arg Ser His Asn Ser Ile Asn Gln Ser Leu Asp Glu Ser Val
            380                 385                 390
Ile Tyr Gly Pro Gly Asn Leu Ser Thr Asn Asn Asn Thr Leu Arg
            395                 400                 405
Asp His Phe Lys Gly Ile Ser Ser His Ser Arg Ser Ser Leu Met
            410                 415                 420
Pro Leu Arg Asn Asp Val Asp Lys Arg Gly Glu Thr Thr Ser Ala
            425                 430                 435
Ser Leu Leu Asn Ala Gly Leu Ser His Thr Glu Tyr Gln Ile Val
            440                 445                 450
Thr Asp Pro Gly Asp Val Pro Ala Phe Leu Lys Trp Leu Ser Leu
            455                 460                 465
Ala Ser Leu Leu Val Tyr Val Ala Ala Phe Ser Ile Gly Leu Gly
            470                 475                 480
Pro Met Pro Trp Leu Val Leu Ser Glu Ile Phe Pro Gly Gly Ile
```

```
                485                 490                 495
Arg Gly Arg Ala Met Ala Leu Thr Ser Ser Met Asn Trp Gly Ile
                500                 505                 510
Asn Leu Leu Ile Ser Leu Thr Phe Leu Thr Val Thr Asp Leu Ile
                515                 520                 525
Gly Leu Pro Trp Val Cys Phe Ile Tyr Thr Ile Met Ser Leu Ala
                530                 535                 540
Ser Leu Leu Phe Val Val Met Pho Ile Pro Glu Thr Lys Gly Cys
                545                 550                 555
Ser Leu Glu Gln Ile Ser Met Glu Leu Ala Lys Val Asn Tyr Val
                560                 565                 570
Lys Asn Asn Ile Cys Phe Met Ser His His Gln Glu Glu Leu Val
                575                 580                 585
Pro Lys Gln Pro Gln Lys Arg Lys Pro Gln Glu Gln Leu Leu Glu
                590                 595                 600
Cys Asn Lys Leu Cys Gly Arg Gly Gln Ser Arg Gln Leu Ser Pro
                605                 610                 615
Glu Thr
```

<210> 2
<211> 1193
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55138203CD1

<400> 2

```
Met Tyr Ser Ala Asn Ile Gly Tyr Leu Leu Phe Val Gly Thr Gly
  1               5                  10                  15
Val Glu Lys Met Asn Asn Thr Pro Ser Met Ala Leu Gly Ser Ser
                20                  25                  30
His Ser Gly Arg Gly Asn Leu Thr Gln Ala Ala Thr Lys Pro Ser
                35                  40                  45
Gly Tyr Glu Lys Thr Asp Asp Val Ser Glu Lys Thr Ser Leu Ala
                50                  55                  60
Asp Gln Glu Glu Val Arg Thr Ile Phe Ile Asn Gln Pro Gln Leu
                65                  70                  75
Thr Lys Phe Cys Asn Asn His Val Ser Thr Ala Lys Tyr Asn Ile
                80                  85                  90
Ile Thr Phe Leu Pro Arg Phe Leu Tyr Ser Gln Phe Arg Arg Ala
                95                 100                 105
Ala Asn Ser Phe Phe Leu Phe Ile Ala Leu Leu Gln Gln Ile Pro
               110                 115                 120
Asp Val Ser Pro Thr Gly Arg Tyr Thr Thr Leu Val Pro Leu Leu
               125                 130                 135
Phe Ile Leu Ala Val Ala Ala Ile Lys Glu Ile Ile Glu Asp Ile
               140                 145                 150
Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys Lys Gln Thr Gln
               155                 160                 165
Val Leu Arg Asn Gly Ala Trp Glu Ile Val His Trp Glu Lys Val
               170                 175                 180
Asn Val Gly Asp Ile Val Ile Ile Lys Gly Lys Glu Tyr Ile Pro
               185                 190                 195
```



```
Ala Asp Thr Val Leu Leu Ser Ser Ser Glu Pro Gln Ala Met Cys
                200                 205                 210
Tyr Ile Glu Thr Ser Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile
                215                 220                 225
Arg Gln Gly Leu Pro Ala Thr Ser Asp Ile Lys Asp Val Asp Ser
                230                 235                 240
Leu Met Arg Ile Ser Gly Arg Ile Glu Cys Glu Ser Pro Asn Arg
                245                 250                 255
His Leu Tyr Asp Phe Val Gly Asn Ile Arg Leu Asp Gly His Gly
                260                 265                 270
Thr Val Pro Leu Gly Ala Asp Gln Ile Leu Leu Arg Gly Ala Gln
                275                 280                 285
Leu Arg Asn Thr Gln Trp Val His Gly Ile Val Val Tyr Thr Gly
                290                 295                 300
His Asp Thr Lys Leu Met Gln Asn Ser Thr Ser Pro Pro Leu Lys
                305                 310                 315
Leu Ser Asn Val Glu Arg Ile Thr Asn Val Gln Ile Leu Ile Leu
                320                 325                 330
Phe Cys Ile Leu Ile Ala Met Ser Leu Val Cys Ser Val Gly Ser
                335                 340                 345
Ala Ile Trp Asn Arg Arg His Ser Gly Lys Asp Trp Tyr Leu Asn
                350                 355                 360
Leu Asn Tyr Gly Gly Ala Ser Asn Phe Gly Leu Asn Phe Leu Thr
                365                 370                 375
Phe Ile Ile Leu Phe Asn Asn Leu Ile Pro Ile Ser Leu Leu Val
                380                 385                 390
Thr Leu Glu Val Val Lys Phe Thr Gln Ala Tyr Phe Ile Asn Trp
                395                 400                 405
Asp Leu Asp Met His Tyr Glu Pro Thr Asp Thr Ala Ala Met Ala
                410                 415                 420
Arg Thr Ser Asn Leu Asn Glu Glu Leu Gly Gln Val Lys Tyr Ile
                425                 430                 435
Phe Ser Asp Lys Thr Gly Thr Leu Thr Cys Asn Val Met Gln Phe
                440                 445                 450
Lys Lys Cys Thr Ile Ala Gly Val Ala Tyr Gly His Val Pro Glu
                455                 460                 465
Pro Glu Asp Tyr Gly Cys Ser Pro Asp Glu Trp Gln Asn Ser Gln
                470                 475                 480
Phe Gly Asp Glu Lys Thr Phe Ser Asp Ser Ser Leu Leu Glu Asn
                485                 490                 495
Leu Gln Asn Asn His Pro Thr Ala Pro Ile Ile Cys Glu Phe Leu
                500                 505                 510
Thr Met Met Ala Val Cys His Thr Ala Val Pro Glu Arg Glu Gly
                515                 520                 525
Asp Lys Ile Ile Tyr Gln Ala Ala Ser Pro Asp Glu Gly Ala Leu
                530                 535                 540
Val Arg Ala Ala Lys Gln Leu Asn Phe Val Phe Thr Gly Arg Thr
                545                 550                 555
Pro Asp Ser Val Ile Ile Asp Ser Leu Gly Gln Glu Glu Arg Tyr
                560                 565                 570
Glu Leu Leu Asn Val Leu Glu Phe Thr Ser Ala Arg Lys Arg Met
                575                 580                 585
Ser Val Ile Val Arg Thr Pro Ser Gly Lys Leu Arg Leu Tyr Cys
                590                 595                 600
Lys Gly Ala Asp Thr Val Ile Tyr Asp Arg Leu Ala Glu Thr Ser
                605                 610                 615
```

```
Lys Tyr Lys Glu Ile Thr Leu Lys His Leu Glu Gln Phe Ala Thr
                620                 625                 630
Glu Gly Leu Arg Thr Leu Cys Phe Ala Val Ala Glu Ile Ser Glu
                635                 640                 645
Ser Asp Phe Gln Glu Trp Arg Ala Val Tyr Gln Arg Ala Ser Thr
                650                 655                 660
Ser Val Gln Asn Arg Leu Leu Lys Leu Glu Glu Ser Tyr Glu Leu
                665                 670                 675
Ile Glu Lys Asn Leu Gln Leu Leu Gly Ala Thr Ala Ile Glu Asp
                680                 685                 690
Lys Leu Gln Asp Gln Val Pro Glu Thr Ile Glu Thr Leu Met Lys
                695                 700                 705
Ala Asp Ile Lys Ile Trp Ile Leu Thr Gly Asp Lys Gln Glu Thr
                710                 715                 720
Ala Ile Asn Ile Gly His Ser Cys Lys Leu Leu Lys Lys Asn Met
                725                 730                 735
Gly Met Ile Val Ile Asn Glu Gly Ser Leu Asp Gly Thr Arg Glu
                740                 745                 750
Thr Leu Ser Arg His Cys Thr Thr Leu Gly Asp Ala Leu Arg Lys
                755                 760                 765
Glu Asn Asp Phe Ala Leu Ile Ile Asp Gly Lys Thr Leu Lys Tyr
                770                 775                 780
Ala Leu Thr Phe Gly Val Arg Gln Tyr Phe Leu Asp Leu Ala Leu
                785                 790                 795
Ser Cys Lys Ala Val Ile Cys Cys Arg Val Ser Pro Leu Gln Lys
                800                 805                 810
Ser Glu Val Val Glu Met Val Lys Lys Gln Val Lys Val Val Thr
                815                 820                 825
Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Thr
                830                 835                 840
Ala His Val Gly Val Gly Ile Ser Gly Asn Glu Gly Leu Gln Ala
                845                 850                 855
Ala Asn Ser Ser Asp Tyr Ser Ile Ala Gln Phe Lys Tyr Leu Lys
                860                 865                 870
Asn Leu Leu Met Ile His Gly Ala Trp Asn Tyr Asn Arg Val Ser
                875                 880                 885
Lys Cys Ile Leu Tyr Cys Phe Tyr Lys Asn Ile Val Leu Tyr Ile
                890                 895                 900
Ile Glu Ile Trp Phe Ala Phe Val Asn Gly Phe Ser Gly Gln Ile
                905                 910                 915
Leu Phe Glu Arg Trp Cys Ile Gly Leu Tyr Asn Val Met Phe Thr
                920                 925                 930
Ala Met Pro Pro Leu Thr Leu Gly Ile Phe Glu Arg Ser Cys Arg
                935                 940                 945
Lys Glu Asn Met Leu Lys Tyr Pro Glu Leu Tyr Lys Thr Ser Gln
                950                 955                 960
Asn Ala Leu Asp Phe Asn Thr Lys Val Phe Trp Val His Cys Leu
                965                 970                 975
Asn Gly Leu Phe His Ser Val Ile Leu Phe Trp Phe Pro Leu Lys
                980                 985                 990
Ala Leu Gln Tyr Gly Thr Ala Phe Gly Asn Gly Lys Thr Ser Asp
                995                1000                1005
Tyr Leu Leu Leu Gly Asn Phe Val Tyr Thr Phe Val Val Ile Thr
               1010                1015                1020
Val Cys Leu Lys Ala Gly Leu Glu Thr Ser Tyr Trp Thr Trp Phe
               1025                1030                1035
```



```
Ser His Ile Ala Ile Trp Gly Ser Ile Ala Leu Trp Val Val Phe
               1040                1045                1050
Leu Gly Ile Tyr Ser Ser Leu Trp Pro Ala Ile Pro Met Ala Pro
               1055                1060                1065
Asp Met Ser Gly Glu Ala Ala Met Leu Phe Ser Ser Gly Val Phe
               1070                1075                1080
Trp Met Gly Leu Leu Phe Ile Pro Val Ala Ser Leu Leu Leu Asp
               1085                1090                1095
Val Val Tyr Lys Val Ile Lys Arg Thr Ala Phe Lys Thr Leu Val
               1100                1105                1110
Asp Glu Val Gln Glu Leu Glu Ala Lys Ser Gln Asp Pro Gly Ala
               1115                1120                1125
Val Val Leu Gly Lys Ser Leu Thr Glu Arg Ala Gln Leu Leu Lys
               1130                1135                1140
Asn Val Phe Lys Lys Asn His Val Asn Leu Tyr Arg Ser Glu Ser
               1145                1150                1155
Leu Gln Gln Asn Leu Leu His Gly Tyr Ala Phe Ser Gln Asp Glu
               1160                1165                1170
Asn Gly Ile Val Ser Gln Ser Glu Val Ile Arg Ala Tyr Asp Thr
               1175                1180                1185
Thr Lys Gln Arg Pro Asp Glu Trp
               1190
```

<210> 3
<211> 989
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No:. 7478871CD1

<400> 3

```
Met Gln Pro Ala Arg Gly Pro Leu Ala Ser Glu Pro Arg Thr Val
  1               5                  10                  15
Leu Val Leu Arg Phe Cys Ala Ser Leu Met Glu Met Lys Leu Pro
                20                  25                  30
Gly Gln Glu Gly Phe Glu Ala Ser Ser Ala Pro Arg Asn Ile Pro
                35                  40                  45
Ser Gly Glu Leu Asp Ser Asn Pro Asp Pro Gly Thr Gly Pro Ser
                50                  55                  60
Pro Asp Gly Pro Ser Asp Thr Glu Ser Lys Glu Leu Gly Val Pro
                65                  70                  75
Lys Asp Pro Leu Leu Phe Ile Gln Leu Asn Glu Leu Leu Gly Trp
                80                  85                  90
Pro Gln Ala Leu Glu Trp Arg Glu Thr Gly Thr Trp Val Leu Phe
                95                 100                 105
Glu Glu Lys Leu Glu Val Ala Ala Gly Arg Trp Ser Ala Pro His
               110                 115                 120
Val Pro Thr Leu Ala Leu Pro Ser Leu Gln Lys Leu Arg Ser Leu
               125                 130                 135
Leu Ala Glu Gly Leu Val Leu Leu Asp Cys Pro Ala Gln Ser Leu
               140                 145                 150
Leu Glu Leu Val Glu Gln Val Thr Arg Val Glu Ser Leu Ser Pro
               155                 160                 165
Glu Leu Arg Gly Gln Leu Gln Ala Leu Leu Leu Gln Arg Pro Gln
```

```
                170                 175                 180
His Tyr Asn Gln Thr Thr Gly Thr Arg Pro Cys Trp Gly Glu Ser
                185                 190                 195
Pro Ser Leu Gly Pro Gly Pro Arg Pro Cys Thr Thr Arg Pro Gln
                200                 205                 210
Ala Pro Gly Pro Ala Gly Gln Cys Gln Asn Pro Leu Arg Gln Lys
                215                 220                 225
Leu Pro Pro Gly Ala Glu Ala Gly Thr Val Leu Ala Gly Glu Leu
                230                 235                 240
Gly Phe Leu Ala Gln Pro Leu Gly Ala Phe Val Arg Leu Arg Asn
                245                 250                 255
Pro Val Val Leu Gly Ser Leu Thr Glu Val Ser Leu Pro Ser Arg
                260                 265                 270
Pho Phe Cys Leu Leu Leu Gly Pro Cys Met Leu Gly Lys Gly Tyr
                275                 280                 285
His Glu Met Gly Arg Ala Ala Ala Val Leu Leu Ser Asp Pro Gln
                290                 295                 300
Pho Gln Trp Ser Val Arg Arg Ala Ser Asn Leu His Asp Leu Leu
                305                 310                 315
Ala Ala Leu Asp Ala Phe Leu Glu Glu Val Thr Val Leu Pro Pro
                320                 325                 330
Gly Arg Trp Asp Pro Thr Ala Arg Ile Pro Pro Pro Lys Cys Leu
                335                 340                 345
Pro Ser Gln His Lys Arg Leu Pro Ser Gln Gln Arg Glu Ile Arg
                350                 355                 360
Gly Pro Ala Val Pro Arg Leu Thr Ser Ala Glu Asp Arg His Arg
                365                 370                 375
His Gly Pro His Ala His Ser Pro Glu Leu Gln Arg Thr Gly Arg
                380                 385                 390
Leu Phe Gly Gly Leu Ile Gln Asp Val Arg Arg Lys Val Pro Trp
                395                 400                 405
Tyr Pro Ser Asp Phe Leu Asp Ala Leu His Leu Gln Cys Phe Ser
                410                 415                 420
Ala Val Leu Tyr Ile Tyr Leu Ala Thr Val Thr Asn Ala Ile Thr
                425                 430                 435
Pho Gly Gly Leu Leu Gly Asp Ala Thr Asp Gly Ala Gln Gly Val
                440                 445                 450
Leu Glu Ser Phe Leu Gly Thr Ala Val Ala Gly Ala Ala Phe Cys
                455                 460                 465
Leu Met Ala Gly Gln Pro Leu Thr Ile Leu Ser Ser Thr Gly Pro
                470                 475                 480
Val Leu Val Phe Glu Arg Leu Leu Phe Ser Phe Ser Arg Asp Tyr
                485                 490                 495
Ser Leu Asp Tyr Leu Pro Phe Arg Leu Trp Val Gly Ile Trp Val
                500                 505                 510
Ala Thr Phe Cys Leu Val Leu Val Ala Thr Glu Ala Ser Val Leu
                515                 520                 525
Val Arg Tyr Phe Thr Arg Phe Thr Glu Glu Gly Phe Cys Ala Leu
                530                 535                 540
Ile Ser Leu Ile Phe Ile Tyr Asp Ala Val Gly Lys Met Leu Asn
                545                 550                 555
Leu Thr His Thr Tyr Pro Ile Gln Lys Pro Gly Ser Ser Ala Tyr
                560                 565                 570
Gly Cys Leu Cys Gln Tyr Pro Gly Pro Gly Gly Asn Glu Ser Gln
                575                 580                 585
Trp Ile Arg Thr Arg Pro Lys Asp Arg Asp Asp Ile Val Ser Met
```



```
                590                 595                 600
Asp Leu Gly Leu Ile Asn Ala Ser Leu Leu Pro Pro Pro Glu Cys
                605                 610                 615
Thr Arg Gln Gly Gly His Pro Arg Gly Pro Gly Cys His Thr Val
                620                 625                 630
Pro Asp Ile Ala Phe Phe Ser Leu Leu Leu Phe Leu Thr Ser Phe
                635                 640                 645
Phe Phe Ala Met Ala Leu Lys Cys Val Lys Thr Ser Arg Phe Phe
                650                 655                 660
Pro Ser Val Val Arg Lys Gly Leu Ser Asp Phe Ser Ser Val Leu
                665                 670                 675
Ala Ile Leu Leu Gly Cys Gly Leu Asp Ala Phe Leu Gly Leu Ala
                680                 685                 690
Thr Pro Lys Leu Met Val Pro Arg Glu Phe Lys Pro Thr Leu Pro
                695                 700                 705
Gly Arg Gly Trp Leu Val Ser Pro Phe Gly Ala Asn Pro Trp Trp
                710                 715                 720
Trp Ser Val Ala Ala Ala Leu Pro Ala Leu Leu Leu Ser Ile Leu
                725                 730                 735
Ile Phe Met Asp Gln Gln Ile Thr Ala Val Ile Leu Asn Arg Met
                740                 745                 750
Glu Tyr Arg Leu Gln Lys Gly Ala Gly Phe His Leu Asp Leu Phe
                755                 760                 765
Cys Val Ala Val Leu Met Leu Leu Thr Ser Ala Leu Gly Leu Pro
                770                 775                 780
Trp Tyr Val Ser Ala Thr Val Ile Ser Leu Ala His Met Asp Ser
                785                 790                 795
Leu Arg Arg Glu Ser Arg Ala Cys Ala Pro Gly Glu Arg Pro Asn
                800                 805                 810
Phe Leu Gly Ile Arg Glu Gln Arg Leu Thr Gly Leu Val Val Phe
                815                 820                 825
Ile Leu Thr Gly Ala Ser Ile Phe Leu Ala Pro Val Leu Lys Phe
                830                 835                 840
Ile Pro Met Pro Val Leu Tyr Gly Ile Phe Leu Tyr Met Gly Val
                845                 850                 855
Ala Ala Leu Ser Ser Ile Gln Phe Thr Asn Arg Val Lys Leu Leu
                860                 865                 870
Leu Met Pro Ala Lys His Gln Pro Asp Leu Leu Leu Leu Arg His
                875                 880                 885
Val Pro Leu Thr Arg Val His Leu Phe Thr Ala Ile Gln Leu Ala
                890                 895                 900
Cys Leu Gly Leu Leu Trp Ile Ile Lys Ser Thr Pro Ala Ala Ile
                905                 910                 915
Ile Phe Pro Leu Met Leu Leu Gly Leu Val Gly Val Arg Lys Ala
                920                 925                 930
Leu Glu Arg Val Phe Ser Pro Gln Glu Leu Leu Trp Leu Asp Glu
                935                 940                 945
Leu Met Pro Glu Glu Glu Arg Ser Ile Pro Glu Lys Gly Leu Glu
                950                 955                 960
Pro Glu His Ser Phe Ser Gly Ser Asp Ser Glu Asp Ser Glu Leu
                965                 970                 975
Met Tyr Gln Pro Lys Ala Pro Glu Ile Asn Ile Ser Val Asn
                980                 985
```

<210> 4
<211> 505

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483601CD1

<400> 4
Met Asp His Ala Glu Glu Asn Glu Ile Leu Ala Ala Thr Gln Arg
1 5 10 15
Tyr Tyr Val Glu Arg Pro Ile Phe Ser His Pro Val Leu Gln Glu
20 25 30
Arg Leu His Thr Lys Asp Lys Val Pro Asp Ser Ile Ala Asp Lys
35 40 45
Leu Lys Gln Ala Phe Thr Cys Thr Pro Lys Lys Ile Arg Asn Ile
50 55 60
Ile Tyr Met Phe Leu Pro Ile Thr Lys Trp Leu Pro Ala Tyr Lys
65 70 75
Phe Lys Glu Tyr Val Leu Gly Asp Leu Val Ser Gly Ile Ser Thr
80 85 90
Gly Val Leu Gln Leu Pro Gln Gly Leu Ala Phe Ala Met Leu Ala
95 100 105
Ala Val Pro Pro Ile Phe Gly Leu Tyr Pro Ser Phe Tyr Pro Val
110 115 120
Ile Met Tyr Cys Phe Leu Gly Thr Ser Arg His Ile Ser Ile Gly
125 130 135
Pro Phe Ala Val Ile Ser Leu Met Ile Gly Gly Val Ala Val Arg
140 145 150
Leu Val Pro Asp Asp Ile Val Ile Pro Gly Gly Val Asn Ala Thr
155 160 165
Asn Gly Thr Glu Ala Arg Asp Ala Leu Arg Val Lys Val Ala Met
170 175 180
Ser Val Thr Leu Leu Ser Gly Ile Ile Gln Phe Cys Leu Gly Val
185 190 195
Cys Arg Phe Gly Phe Val Ala Ile Tyr Leu Thr Glu Pro Leu Val
200 205 210
Arg Gly Phe Thr Thr Ala Ala Ala Val His Val Phe Thr Ser Met
215 220 225
Leu Lys Tyr Leu Phe Gly Val Lys Thr Lys Arg Tyr Ser Gly Ile
230 235 240
Phe Ser Val Val Tyr Ser Thr Val Ala Val Leu Gln Asn Val Lys
245 250 255
Asn Leu Asn Val Cys Ser Leu Gly Val Gly Leu Met Val Phe Gly
260 265 270
Leu Leu Leu Gly Gly Lys Glu Phe Asn Glu Arg Phe Lys Glu Lys
275 280 285
Leu Pro Ala Pro Ile Pro Leu Glu Phe Phe Ala Val Val Met Gly
290 295 300
Thr Gly Ile Ser Ala Gly Phe Asn Leu Lys Glu Ser Tyr Asn Val
305 310 315
Asp Val Val Gly Thr Leu Pro Leu Gly Leu Leu Pro Pro Ala Asn
320 325 330
Pro Asp Thr Ser Leu Phe His Leu Val Tyr Val Asp Ala Ile Ala
335 340 345
Ile Ala Ile Val Gly Phe Ser Val Thr Ile Ser Met Ala Lys Thr
350 355 360

Leu Ala Asn Lys His Gly Tyr Gln Val Asp Gly Asn Gln Glu Leu
365 370 375
Ile Ala Leu Gly Leu Cys Asn Ser Ile Gly Ser Leu Phe Gln Thr
380 385 390
Phe Ser Ile Ser Cys Ser Leu Ser Arg Ser Leu Val Gln Glu Gly
395 400 405
Thr Gly Gly Lys Thr Gln Leu Ala Gly Cys Leu Ala Ser Leu Met
410 415 420
Ile Leu Leu Val Ile Leu Ala Thr Gly Phe Leu Phe Glu Ser Leu
425 430 435
Pro Gln Ala Val Leu Ser Ala Ile Val Ile Val Asn Leu Lys Gly
440 445 450
Met Phe Met Gln Phe Ser Asp Leu Pro Phe Phe Trp Arg Thr Ser
455 460 465
Lys Ile Glu Leu Thr Ile Trp Leu Thr Thr Phe Val Ser Ser Leu
470 475 480
Phe Leu Gly Leu Asp Tyr Gly Leu Ile Thr Ala Val Ile Ile Ala
485 490 495
Leu Leu Thr Val Ile Tyr Arg Thr Gln Arg
500 505

<210> 5
<211> 618
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487851CD1

<400> 5
Met Ser Arg Ser Pro Leu Asn Pro Ser Gln Leu Arg Ser Val Gly
1 5 10 15
Ser Gln Asp Ala Leu Ala Pro Leu Pro Pro Pro Ala Pro Gln Asn
20 25 30
Pro Ser Thr His Ser Trp Asp Pro Leu Cys Gly Ser Leu Pro Trp
35 40 45
Gly Leu Ser Cys Leu Leu Ala Leu Gln His Val Leu Val Met Ala
50 55 60
Ser Leu Leu Cys Val Ser His Leu Leu Leu Leu Cys Ser Leu Ser
65 70 75
Pro Gly Gly Leu Ser Tyr Ser Pro Ser Gln Leu Leu Ala Ser Ser
80 85 90
Phe Phe Ser Arg Gly Met Ser Thr Ile Leu Gln Thr Trp Met Gly
95 100 105
Ser Arg Leu Pro Leu Val Gln Ala Pro Ser Leu Glu Phe Leu Ile
110 115 120
Pro Ala Leu Val Leu Thr Ser Gln Lys Leu Pro Arg Ala Ile Gln
125 130 135
Thr Pro Gly Asn Cys Glu His Arg Ala Arg Ala Arg Ala Ser Leu
140 145 150
Met Leu His Leu Cys Arg Gly Pro Ser Cys His Gly Leu Gly His
155 160 165
Trp Asn Thr Ser Leu Gln Glu Val Ser Gly Ala Val Val Val Ser
170 175 180
Gly Leu Leu Gln Gly Met Met Gly Leu Leu Gly Ser Pro Gly His

```
                185                 190                 195
Val Phe Pro His Cys Gly Pro Leu Val Leu Ala Pro Ser Leu Val
                200                 205                 210
Val Ala Gly Leu Ser Ala His Arg Glu Val Ala Gln Phe Cys Phe
                215                 220                 225
Thr His Trp Gly Leu Ala Leu Leu Val Ile Leu Leu Met Val Val
                230                 235                 240
Cys Ser Gln His Leu Gly Ser Cys Gln Phe His Val Cys Pro Trp
                245                 250                 255
Arg Arg Ala Ser Thr Ser Ser Thr His Thr Pro Leu Pro Val Phe
                260                 265                 270
Arg Leu Leu Ser Val Leu Ile Pro Val Ala Cys Val Trp Ile Val
                275                 280                 285
Ser Ala Phe Val Gly Phe Ser Val Ile Pro Gln Glu Leu Ser Ala
                290                 295                 300
Pro Thr Lys Ala Pro Trp Ile Trp Leu Pro His Pro Gly Glu Trp
                305                 310                 315
Asn Trp Pro Leu Leu Thr Pro Arg Ala Leu Ala Ala Gly Ile Ser
                320                 325                 330
Met Ala Leu Ala Ala Ser Thr Ser Ser Leu Gly Cys Tyr Ala Leu
                335                 340                 345
Cys Gly Arg Leu Leu His Leu Pro Pro Pro Pro Pro His Ala Cys
                350                 355                 360
Ser Arg Gly Leu Ser Leu Glu Gly Leu Gly Ser Val Leu Ala Gly
                365                 370                 375
Leu Leu Gly Ser Pro Met Gly Thr Ala Ser Ser Phe Pro Asn Val
                380                 385                 390
Gly Lys Val Gly Leu Ile Gln Ala Gly Ser Gln Gln Val Ala His
                395                 400                 405
Leu Val Gly Leu Leu Cys Val Gly Leu Gly Leu Ser Pro Arg Leu
                410                 415                 420
Ala Gln Leu Leu Thr Thr Ile Pro Leu Pro Val Val Gly Gly Val
                425                 430                 435
Leu Gly Val Thr Gln Ala Val Val Leu Ser Ala Gly Phe Ser Ser
                440                 445                 450
Phe Tyr Leu Ala Asp Ile Asp Ser Gly Arg Asn Ile Phe Ile Val
                455                 460                 465
Gly Phe Ser Ile Phe Met Ala Leu Leu Leu Pro Arg Trp Phe Arg
                470                 475                 480
Glu Ala Pro Val Leu Phe Ser Thr Gly Trp Ser Pro Leu Asp Val
                485                 490                 495
Leu Leu His Ser Leu Leu Thr Gln Pro Ile Phe Leu Ala Gly Leu
                500                 505                 510
Ser Gly Phe Leu Leu Glu Asn Thr Ile Pro Gly Thr Gln Leu Glu
                515                 520                 525
Arg Gly Leu Gly Gln Gly Leu Pro Ser Pro Phe Thr Ala Gln Glu
                530                 535                 540
Ala Arg Met Pro Gln Lys Pro Arg Glu Lys Ala Ala Gln Val Tyr
                545                 550                 555
Arg Leu Pro Phe Pro Ile Gln Asn Leu Cys Pro Cys Ile Pro Gln
                560                 565                 570
Pro Leu His Cys Leu Cys Pro Leu Pro Glu Asp Pro Gly Asp Glu
                575                 580                 585
Glu Gly Gly Ser Ser Glu Pro Glu Glu Met Ala Asp Leu Leu Pro
                590                 595                 600
Gly Ser Gly Glu Pro Cys Pro Glu Ser Ser Arg Glu Gly Phe Arg
```



```
                605                 610                 615
Ser Gln Lys
```

<210> 6
<211> 377
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472881CD1

<400> 6

```
Met Arg Ala Asn Cys Ser Ser Ser Ser Ala Cys Pro Ala Asn Ser
  1               5                  10                  15
Ser Glu Glu Glu Leu Pro Val Gly Leu Glu Ala His Gly Asn Leu
                20                  25                  30
Glu Leu Val Phe Thr Val Val Pro Thr Val Met Met Gly Leu Leu
                35                  40                  45
Met Phe Ser Leu Gly Cys Ser Val Glu Ile Arg Lys Leu Trp Ser
                50                  55                  60
His Ile Arg Arg Pro Trp Gly Ile Ala Val Gly Leu Leu Cys Gln
                65                  70                  75
Phe Gly Leu Met Pro Phe Thr Ala Tyr Leu Leu Ala Ile Ser Phe
                80                  85                  90
Ser Leu Lys Pro Val Gln Ala Ile Ala Val Leu Ile Met Gly Cys
                95                 100                 105
Cys Pro Gly Gly Thr Ile Ser Asn Ile Phe Thr Phe Trp Val Asp
               110                 115                 120
Gly Asp Met Asp Leu Ser Ile Ser Met Thr Thr Cys Ser Thr Val
               125                 130                 135
Ala Ala Leu Gly Met Met Pro Leu Cys Ile Tyr Leu Tyr Thr Trp
               140                 145                 150
Ser Trp Ser Leu Gln Gln Asn Leu Thr Ile Pro Tyr Gln Asn Ile
               155                 160                 165
Gly Ile Thr Leu Val Cys Leu Thr Ile Pro Val Ala Phe Gly Val
               170                 175                 180
Tyr Val Asn Tyr Arg Trp Pro Lys Gln Ser Lys Ile Ile Leu Lys
               185                 190                 195
Ile Gly Ala Val Val Gly Gly Val Leu Leu Leu Val Val Ala Val
               200                 205                 210
Ala Gly Val Val Leu Ala Lys Gly Ser Trp Asn Ser Asp Ile Thr
               215                 220                 225
Leu Leu Thr Ile Ser Phe Ile Phe Pro Leu Ile Gly His Val Thr
               230                 235                 240
Gly Phe Leu Leu Ala Leu Phe Thr His Gln Ser Trp Gln Arg Cys
               245                 250                 255
Arg Thr Ile Ser Leu Glu Thr Gly Ala Gln Asn Ile Gln Met Cys
               260                 265                 270
Ile Thr Met Leu Gln Leu Ser Phe Thr Ala Glu His Leu Val Gln
               275                 280                 285
Met Leu Ser Phe Pro Leu Ala Tyr Gly Leu Phe Gln Leu Ile Asp
               290                 295                 300
Gly Phe Leu Ile Val Ala Ala Tyr Gln Thr Tyr Lys Arg Arg Leu
               305                 310                 315
```

```
Lys Asn Lys His Gly Lys Lys Asn Ser Gly Cys Thr Glu Val Cys
                320                 325                 330
His Thr Arg Lys Ser Thr Ser Ser Arg Glu Thr Asn Ala Phe Leu
                335                 340                 345
Glu Val Asn Glu Glu Gly Ala Ile Thr Pro Gly Pro Pro Gly Pro
                350                 355                 360
Met Asp Cys His Arg Ala Leu Glu Pro Val Gly His Ile Thr Ser
                365                 370                 375
Cys Glu
```

<210> 7
<211> 507
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7612560CD1

<400> 7

```
Met Ser Val Thr Lys Ser Thr Glu Gly Pro Gln Gly Ala Val Ala
  1               5                  10                  15
Ile Lys Leu Asp Leu Met Ser Pro Pro Glu Ser Ala Lys Lys Leu
                20                  25                  30
Glu Asn Lys Asp Ser Thr Phe Leu Asp Glu Ser Pro Ser Glu Ser
                35                  40                  45
Ala Gly Leu Lys Lys Thr Lys Gly Ile Thr Val Phe Gln Ala Leu
                50                  55                  60
Ile His Leu Val Lys Gly Asn Met Gly Thr Gly Ile Leu Gly Leu
                65                  70                  75
Pro Leu Ala Val Lys Asn Ala Gly Ile Leu Met Gly Pro Leu Ser
                80                  85                  90
Leu Leu Val Met Gly Phe Ile Ala Cys His Cys Met His Ile Leu
                95                 100                 105
Val Lys Cys Ala Gln Arg Phe Cys Lys Arg Leu Asn Lys Pro Phe
               110                 115                 120
Met Asp Tyr Gly Asp Thr Val Met His Gly Leu Glu Ala Asn Pro
               125                 130                 135
Asn Ala Trp Leu Gln Asn His Ala His Trp Gly Arg His Ile Val
               140                 145                 150
Ser Phe Phe Leu Ile Ile Thr Gln Leu Gly Phe Cys Cys Val Tyr
               155                 160                 165
Ile Val Phe Leu Ala Asp Asn Leu Lys Gln Val Val Glu Ala Val
               170                 175                 180
Asn Ser Thr Thr Asn Asn Cys Tyr Ser Asn Glu Thr Val Ile Leu
               185                 190                 195
Thr Pro Thr Met Asp Ser Arg Leu Tyr Met Leu Ser Phe Leu Pro
               200                 205                 210
Phe Leu Val Leu Leu Val Leu Ile Arg Asn Leu Arg Ile Leu Thr
               215                 220                 225
Ile Phe Ser Met Leu Ala Asn Ile Ser Met Leu Val Ser Leu Val
               230                 235                 240
Ile Ile Ile Gln Tyr Ile Thr Gln Glu Ile Pro Asp Pro Ser Arg
               245                 250                 255
Leu Pro Leu Val Ala Ser Trp Lys Thr Tyr Pro Leu Phe Phe Gly
```



```
               260                 265                 270
Thr Ala Ile Phe Ser Phe Glu Ser Ile Gly Val Val Leu Pro Leu
               275                 280                 285
Glu Asn Lys Met Lys Asn Ala Arg His Phe Pro Ala Ile Leu Ser
               290                 295                 300
Leu Gly Met Ser Ile Val Thr Ser Leu Tyr Ile Gly Met Ala Ala
               305                 310                 315
Leu Gly Tyr Leu Arg Phe Gly Asp Asp Ile Lys Ala Ser Ile Ser
               320                 325                 330
Leu Asn Leu Pro Asn Cys Trp Leu Tyr Gln Ser Val Lys Leu Leu
               335                 340                 345
Tyr Ile Ala Gly Ile Leu Cys Thr Tyr Ala Leu Gln Phe Tyr Val
               350                 355                 360
Pro Ala Glu Ile Ile Ile Pro Phe Ala Ile Ser Arg Val Ser Thr
               365                 370                 375
Arg Trp Ala Leu Pro Leu Asp Leu Ser Ile Arg Leu Val Met Val
               380                 385                 390
Cys Leu Thr Cys Leu Leu Ala Ile Leu Ile Pro Arg Leu Asp Leu
               395                 400                 405
Val Ile Ser Leu Val Gly Ser Val Ser Gly Thr Ala Leu Ala Leu
               410                 415                 420
Ile Ile Pro Pro Leu Leu Glu Val Thr Thr Phe Tyr Ser Glu Gly
               425                 430                 435
Met Ser Pro Leu Thr Ile Phe Lys Asp Val Leu Ile Ser Ile Leu
               440                 445                 450
Gly Phe Val Gly Phe Val Val Gly Thr Tyr Gln Ala Leu Asp Glu
               455                 460                 465
Leu Leu Lys Ser Glu Asp Ser His Pro Phe Ser Asn Ser Thr Thr
               470                 475                 480
Phe Val Arg Val Glu Leu Cys Lys Lys Gln Pro Pro Glu Gly Pro
               485                 490                 495
Lys Trp Gln Gln Leu Ala Lys Gly Asp Ala Ala Ser
               500                 505
```

<210> 8
<211> 438
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2880370CD1

<400> 8

```
Met Ile Arg Lys Leu Phe Ile Val Leu Leu Leu Leu Leu Val Thr
  1               5                  10                  15
Ile Glu Glu Ala Arg Met Ser Ser Leu Ser Phe Leu Asn Ile Glu
                20                  25                  30
Lys Thr Glu Ile Leu Phe Phe Thr Lys Thr Glu Glu Thr Ile Leu
                35                  40                  45
Val Ser Ser Ser Tyr Glu Asn Lys Arg Pro Asn Ser Ser His Leu
                50                  55                  60
Phe Val Lys Ile Glu Asp Pro Lys Ile Leu Gln Met Val Asn Val
                65                  70                  75
Ala Lys Lys Ile Ser Ser Asp Ala Thr Asn Phe Thr Ile Asn Leu
                80                  85                  90
```

```
Val Thr Asp Glu Glu Gly Glu Thr Asn Val Thr Ile Gln Leu Trp
                 95                  100                 105
Asp Ser Glu Gly Arg Gln Glu Arg Leu Ile Glu Glu Ile Lys Asn
                110                 115                 120
Val Lys Val Lys Val Leu Lys Gln Lys Asp Ser Leu Leu Gln Ala
                125                 130                 135
Pro Met His Ile Asp Arg Asn Ile Leu Met Leu Ile Leu Pro Leu
                140                 145                 150
Ile Leu Leu Asn Lys Cys Ala Phe Gly Cys Lys Ile Glu Leu Gln
                155                 160                 165
Leu Phe Gln Thr Val Trp Lys Arg Pro Leu Pro Val Ile Leu Gly
                170                 175                 180
Ala Val Thr Gln Phe Phe Leu Met Pro Phe Cys Gly Phe Leu Leu
                185                 190                 195
Ser Gln Ile Val Ala Leu Pro Glu Ala Gln Ala Phe Gly Val Val
                200                 205                 210
Met Thr Cys Thr Cys Pro Gly Gly Gly Gly Gly Tyr Leu Phe Ala
                215                 220                 225
Leu Leu Leu Asp Gly Asp Phe Thr Leu Ala Ile Leu Met Thr Cys
                230                 235                 240
Thr Ser Thr Leu Leu Ala Leu Ile Met Met Pro Val Asn Ser Tyr
                245                 250                 255
Ile Tyr Ser Arg Ile Leu Gly Leu Ser Gly Thr Phe His Ile Pro
                260                 265                 270
Val Ser Lys Ile Val Ser Thr Leu Leu Phe Ile Leu Val Pro Val
                275                 280                 285
Ser Ile Gly Ile Val Ile Lys His Arg Ile Pro Glu Lys Ala Ser
                290                 295                 300
Phe Leu Glu Arg Ile Ile Arg Pro Leu Ser Phe Ile Leu Met Phe
                305                 310                 315
Val Gly Ile Tyr Leu Thr Phe Thr Val Gly Leu Val Phe Leu Lys
                320                 325                 330
Thr Asp Asn Leu Glu Val Ile Leu Leu Gly Leu Leu Val Pro Ala
                335                 340                 345
Leu Gly Leu Leu Phe Gly Tyr Ser Phe Ala Lys Val Cys Thr Leu
                350                 355                 360
Pro Leu Pro Val Cys Lys Thr Val Ala Ile Glu Ser Gly Met Leu
                365                 370                 375
Asn Ser Phe Leu Ala Leu Ala Val Ile Gln Leu Ser Phe Pro Gln
                380                 385                 390
Ser Lys Ala Asn Leu Ala Ser Val Ala Pro Phe Thr Val Ala Met
                395                 400                 405
Cys Ser Gly Cys Glu Met Leu Leu Ile Ile Leu Val Tyr Lys Ala
                410                 415                 420
Lys Lys Arg Cys Ile Phe Phe Leu Gln Asp Lys Arg Lys Arg Asn
                425                 430                 435
Phe Leu Ile
```

<210> 9
<211> 350
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 6267489CD1

<400> 9

```
Met Leu Glu Gly Ala Glu Leu Tyr Phe Asn Val Asp His Gly Tyr
  1               5                  10                  15
Leu Glu Gly Leu Val Arg Gly Cys Lys Ala Ser Leu Leu Thr Gln
                 20                  25                  30
Gln Asp Tyr Ile Asn Leu Val Gln Cys Glu Thr Leu Glu Asp Leu
                 35                  40                  45
Lys Ile His Leu Gln Thr Thr Asp Tyr Gly Asn Phe Leu Ala Asn
                 50                  55                  60
His Thr Asn Pro Leu Thr Val Ser Lys Ile Asp Thr Glu Met Arg
                 65                  70                  75
Lys Arg Leu Cys Gly Glu Phe Glu Tyr Phe Arg Asn His Ser Leu
                 80                  85                  90
Glu Pro Leu Ser Thr Phe Leu Thr Tyr Met Thr Cys Ser Tyr Met
                 95                 100                 105
Ile Asp Asn Val Ile Leu Leu Met Asn Gly Ala Leu Gln Lys Lys
                110                 115                 120
Ser Val Lys Glu Ile Leu Gly Lys Cys His Pro Leu Gly Arg Phe
                125                 130                 135
Thr Glu Met Glu Ala Val Asn Ile Ala Glu Thr Pro Ser Asp Leu
                140                 145                 150
Phe Asn Ala Ile Leu Ile Glu Thr Pro Leu Ala Pro Phe Phe Gln
                155                 160                 165
Asp Cys Met Ser Glu Asn Ala Leu Asp Glu Leu Asn Ile Glu Leu
                170                 175                 180
Leu Arg Asn Lys Leu Tyr Lys Ser Tyr Leu Glu Ala Phe Tyr Lys
                185                 190                 195
Phe Cys Lys Asn His Gly Asp Val Thr Ala Glu Val Met Cys Pro
                200                 205                 210
Ile Leu Glu Phe Glu Ala Asp Arg Arg Ala Phe Ile Ile Thr Leu
                215                 220                 225
Asn Ser Phe Gly Thr Glu Leu Ser Lys Glu Asp Arg Glu Thr Leu
                230                 235                 240
Tyr Pro Thr Phe Gly Lys Leu Tyr Pro Glu Gly Leu Arg Leu Leu
                245                 250                 255
Ala Gln Ala Glu Asp Phe Asp Gln Met Lys Asn Val Ala Asp His
                260                 265                 270
Tyr Gly Val Tyr Lys Pro Leu Phe Glu Ala Val Gly Gly Ser Gly
                275                 280                 285
Gly Lys Thr Leu Glu Asp Val Phe Tyr Glu Arg Glu Val Gln Met
                290                 295                 300
Asn Val Leu Ala Phe Asn Arg Gln Phe His Tyr Gly Val Phe Tyr
                305                 310                 315
Ala Tyr Val Lys Leu Lys Glu Gln Glu Ile Arg Asn Ile Val Trp
                320                 325                 330
Ile Ala Glu Cys Ile Ser Gln Arg His Arg Thr Lys Ile Asn Ser
                335                 340                 345
Tyr Ile Pro Ile Leu
                350
```

<210> 10
<211> 1707
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484777CD1

<400> 10

```
Met Pro Glu Pro Trp Gly Thr Val Tyr Phe Leu Gly Ile Ala Gln
  1               5                  10                  15
Val Phe Ser Phe Leu Phe Ser Trp Trp Asn Leu Glu Gly Val Met
                 20                  25                  30
Asn Gln Ala Asp Ala Pro Arg Pro Leu Asn Trp Thr Ile Arg Lys
                 35                  40                  45
Leu Cys His Ala Ala Phe Leu Pro Ser Val Arg Leu Leu Lys Ala
                 50                  55                  60
Gln Lys Ser Trp Ile Glu Arg Ala Phe Tyr Lys Arg Glu Cys Val
                 65                  70                  75
His Ile Ile Pro Ser Thr Lys Asp Pro His Arg Cys Cys Cys Gly
                 80                  85                  90
Arg Leu Ile Gly Gln His Val Gly Leu Thr Pro Ser Ile Ser Val
                 95                 100                 105
Leu Gln Asn Glu Lys Asn Glu Ser Arg Leu Ser Arg Asn Asp Ile
                110                 115                 120
Gln Ser Glu Lys Trp Ser Ile Ser Lys His Thr Gln Leu Ser Pro
                125                 130                 135
Thr Asp Ala Phe Gly Thr Ile Glu Phe Gln Gly Gly Gly His Ser
                140                 145                 150
Asn Lys Ala Met Tyr Val Arg Val Ser Phe Asp Thr Lys Pro Asp
                155                 160                 165
Leu Leu Leu His Leu Met Thr Lys Glu Trp Gln Leu Glu Leu Pro
                170                 175                 180
Lys Leu Leu Ile Ser Val His Gly Gly Leu Gln Asn Phe Glu Leu
                185                 190                 195
Gln Pro Lys Leu Lys Gln Val Phe Gly Lys Gly Leu Ile Lys Ala
                200                 205                 210
Ala Met Thr Thr Gly Ala Trp Ile Phe Thr Gly Gly Val Asn Thr
                215                 220                 225
Gly Val Ile Arg His Val Gly Asp Ala Leu Lys Asp His Ala Ser
                230                 235                 240
Lys Ser Arg Gly Lys Ile Cys Thr Ile Gly Ile Ala Pro Trp Gly
                245                 250                 255
Ile Val Glu Asn Gln Glu Asp Leu Ile Gly Arg Asp Val Val Arg
                260                 265                 270
Pro Tyr Gln Thr Met Ser Asn Pro Met Ser Lys Leu Thr Val Leu
                275                 280                 285
Asn Ser Met His Ser His Phe Ile Leu Ala Asp Asn Gly Thr Thr
                290                 295                 300
Gly Lys Tyr Gly Ala Glu Val Lys Leu Arg Arg Gln Leu Glu Lys
                305                 310                 315
His Ile Ser Leu Gln Lys Ile Asn Thr Arg Ile Gly Gln Gly Val
                320                 325                 330
Pro Val Val Ala Leu Ile Val Glu Gly Gly Pro Asn Val Ile Ser
                335                 340                 345
Ile Val Leu Glu Tyr Leu Arg Asp Thr Pro Pro Val Pro Val Val
                350                 355                 360
Val Cys Asp Gly Ser Gly Arg Ala Ser Asp Ile Leu Ala Phe Gly
                365                 370                 375
His Lys Tyr Ser Glu Glu Gly Gly Leu Ile Asn Glu Ser Leu Arg
```



```
                380                 385                 390
Asp Gln Leu Leu Val Thr Ile Gln Lys Thr Phe Thr Tyr Thr Arg
                395                 400                 405
Thr Gln Ala Gln His Leu Phe Ile Ile Leu Met Glu Cys Met Lys
                410                 415                 420
Lys Lys Glu Leu Ile Thr Val Phe Arg Met Gly Ser Glu Gly His
                425                 430                 435
Gln Asp Ile Asp Leu Ala Ile Leu Thr Ala Leu Leu Lys Gly Ala
                440                 445                 450
Asn Ala Ser Ala Pro Asp Gln Leu Ser Leu Ala Leu Ala Trp Asn
                455                 460                 465
Arg Val Asp Ile Ala Arg Ser Gln Ile Phe Ile Tyr Gly Gln Gln
                470                 475                 480
Trp Pro Val Gly Ser Leu Glu Gln Ala Met Leu Asp Ala Leu Val
                485                 490                 495
Leu Asp Arg Val Asp Phe Val Lys Leu Leu Ile Glu Asn Gly Val
                500                 505                 510
Ser Met His Arg Phe Leu Thr Ile Ser Arg Leu Glu Glu Leu Tyr
                515                 520                 525
Asn Thr Arg His Gly Pro Ser Asn Thr Leu Tyr His Leu Val Arg
                530                 535                 540
Asp Val Lys Lys Gly Asn Leu Pro Pro Asp Tyr Arg Ile Ser Leu
                545                 550                 555
Ile Asp Ile Gly Leu Val Ile Glu Tyr Leu Met Gly Gly Ala Tyr
                560                 565                 570
Arg Cys Asn Tyr Thr Arg Lys Arg Phe Arg Thr Leu Tyr His Asn
                575                 580                 585
Leu Phe Gly Pro Lys Arg Pro Lys Ala Leu Lys Leu Leu Gly Met
                590                 595                 600
Glu Asp Asp Ile Pro Leu Arg Arg Gly Arg Lys Thr Thr Lys Lys
                605                 610                 615
Arg Glu Glu Glu Val Asp Ile Asp Leu Asp Asp Pro Glu Ile Asn
                620                 625                 630
His Phe Pro Phe Pro Phe His Glu Leu Met Val Trp Ala Val Leu
                635                 640                 645
Met Lys Arg Gln Lys Met Ala Leu Phe Phe Trp Gln His Gly Glu
                650                 655                 660
Glu Ala Met Ala Lys Ala Leu Val Ala Cys Lys Leu Cys Lys Ala
                665                 670                 675
Met Ala His Glu Ala Ser Glu Asn Asp Met Val Asp Asp Ile Ser
                680                 685                 690
Gln Glu Leu Asn His Asn Ser Arg Asp Phe Gly Gln Leu Ala Val
                695                 700                 705
Glu Leu Leu Asp Gln Ser Tyr Lys Gln Asp Glu Gln Leu Ala Met
                710                 715                 720
Lys Leu Leu Thr Tyr Glu Leu Lys Asn Trp Ser Asn Ala Thr Cys
                725                 730                 735
Leu Gln Leu Ala Val Ala Ala Lys His Arg Asp Phe Ile Ala His
                740                 745                 750
Thr Cys Ser Gln Met Leu Leu Thr Asp Met Trp Met Gly Arg Leu
                755                 760                 765
Arg Met Arg Lys Asn Ser Gly Leu Lys Val Ile Leu Gly Ile Leu
                770                 775                 780
Leu Pro Pro Ser Ile Leu Ser Leu Glu Phe Lys Asn Lys Asp Asp
                785                 790                 795
Met Pro Tyr Met Ser Gln Ala Gln Glu Ile His Leu Gln Glu Lys
```

```
                800                 805                 810
Glu Ala Glu Glu Pro Glu Lys Pro Thr Lys Glu Lys Glu Glu Glu
                815                 820                 825
Asp Met Glu Leu Thr Ala Met Leu Gly Arg Asn Asn Gly Glu Ser
                830                 835                 840
Ser Arg Lys Lys Asp Glu Glu Glu Val Gln Ser Lys His Arg Leu
                845                 850                 855
Ile Pro Leu Gly Arg Lys Ile Tyr Glu Phe Tyr Asn Ala Pro Ile
                860                 865                 870
Val Lys Phe Trp Phe Tyr Thr Leu Ala Tyr Ile Gly Tyr Leu Met
                875                 880                 885
Leu Phe Asn Tyr Ile Val Leu Val Lys Met Glu Arg Trp Pro Ser
                890                 895                 900
Thr Gln Glu Trp Ile Val Ile Ser Tyr Ile Phe Thr Leu Gly Ile
                905                 910                 915
Glu Lys Met Arg Glu Ile Leu Met Ser Glu Pro Gly Lys Leu Leu
                920                 925                 930
Gln Lys Val Lys Val Trp Leu Gln Glu Tyr Trp Asn Val Thr Asp
                935                 940                 945
Leu Ile Ala Ile Leu Leu Phe Ser Val Gly Met Ile Leu Arg Leu
                950                 955                 960
Gln Asp Gln Pro Phe Arg Ser Asp Gly Arg Val Ile Tyr Cys Val
                965                 970                 975
Asn Ile Ile Tyr Trp Tyr Ile Arg Leu Leu Asp Ile Phe Gly Val
                980                 985                 990
Asn Lys Tyr Leu Gly Pro Tyr Val Met Met Ile Gly Lys Met Met
                995                1000                1005
Ile Asp Met Met Tyr Phe Val Ile Ile Met Leu Val Val Leu Met
               1010                1015                1020
Ser Phe Gly Val Ala Arg Gln Ala Ile Leu Phe Pro Asn Glu Glu
               1025                1030                1035
Pro Ser Trp Lys Leu Ala Lys Asn Ile Phe Tyr Met Pro Tyr Trp
               1040                1045                1050
Met Ile Tyr Gly Glu Val Phe Ala Asp Gln Ile Asp Pro Pro Cys
               1055                1060                1065
Gly Gln Asn Glu Thr Arg Glu Asp Gly Lys Ile Ile Gln Leu Pro
               1070                1075                1080
Pro Cys Lys Thr Gly Ala Trp Ile Val Pro Ala Ile Met Ala Cys
               1085                1090                1095
Tyr Leu Leu Val Ala Asn Ile Leu Leu Val Asn Leu Leu Ile Ala
               1100                1105                1110
Val Phe Asn Asn Thr Phe Phe Glu Val Lys Ser Ile Ser Asn Gln
               1115                1120                1125
Val Trp Lys Phe Gln Arg Tyr Gln Leu Ile Met Thr Phe His Glu
               1130                1135                1140
Arg Pro Val Leu Pro Pro Pro Leu Ile Ile Phe Ser His Met Thr
               1145                1150                1155
Met Ile Phe Gln His Leu Cys Cys Arg Trp Arg Lys His Glu Ser
               1160                1165                1170
Asp Pro Asp Glu Arg Asp Tyr Gly Leu Lys Leu Phe Ile Thr Asp
               1175                1180                1185
Asp Glu Leu Lys Lys Val His Asp Phe Glu Glu Gln Cys Ile Glu
               1190                1195                1200
Glu Tyr Phe Arg Glu Lys Asp Asp Arg Phe Asn Ser Ser Asn Asp
               1205                1210                1215
Glu Arg Ile Arg Val Thr Ser Glu Arg Val Glu Asn Met Ser Met
```



```
               1220                1225                1230
Arg Leu Glu Glu Val Asn Glu Arg Glu His Ser Met Lys Ala Ser
               1235                1240                1245
Leu Gln Thr Val Asp Ile Arg Leu Ala Gln Leu Glu Asp Leu Ile
               1250                1255                1260
Gly Arg Met Ala Thr Ala Leu Glu Arg Leu Thr Gly Leu Glu Arg
               1265                1270                1275
Ala Glu Ser Asn Lys Ile Arg Ser Arg Thr Ser Ser Asp Cys Thr
               1280                1285                1290
Asp Ala Ala Tyr Ile Val Arg Gln Ser Ser Phe Asn Ser Gln Glu
               1295                1300                1305
Gly Asn Thr Phe Lys Leu Gln Glu Ser Ile Asp Pro Ala Gly Glu
               1310                1315                1320
Glu Thr Met Ser Pro Thr Ser Pro Thr Leu Met Pro Arg Met Arg
               1325                1330                1335
Ser His Ser Phe Tyr Ser Val Asn Met Lys Asp Lys Gly Gly Ile
               1340                1345                1350
Glu Lys Leu Glu Ser Ile Phe Lys Glu Arg Ser Leu Ser Leu His
               1355                1360                1365
Arg Ala Thr Ser Ser His Ser Val Ala Lys Glu Pro Lys Ala Pro
               1370                1375                1380
Ala Ala Pro Ala Asn Thr Leu Ala Ile Val Pro Asp Ser Arg Arg
               1385                1390                1395
Pro Ser Ser Cys Ile Asp Ile Tyr Val Ser Ala Met Asp Glu Leu
               1400                1405                1410
His Cys Asp Ile Asp Pro Leu Asp Asn Ser Val Asn Ile Leu Gly
               1415                1420                1425
Leu Gly Glu Pro Ser Phe Ser Thr Pro Val Pro Ser Thr Ala Pro
               1430                1435                1440
Ser Ser Ser Ala Tyr Ala Thr Leu Ala Pro Thr Asp Arg Pro Pro
               1445                1450                1455
Ser Arg Ser Ile Asp Phe Glu Asp Ile Thr Ser Met Asp Thr Arg
               1460                1465                1470
Ser Phe Ser Ser Asp Tyr Thr His Leu Pro Glu Cys Gln Asn Pro
               1475                1480                1485
Trp Asp Ser Glu Pro Pro Met Tyr His Thr Ile Glu Arg Ser Lys
               1490                1495                1500
Ser Ser Arg Tyr Leu Ala Thr Thr Pro Phe Leu Leu Glu Glu Ala
               1505                1510                1515
Pro Ile Val Lys Ser His Ser Phe Met Phe Ser Pro Ser Arg Ser
               1520                1525                1530
Tyr Tyr Ala Asn Phe Gly Val Pro Val Lys Thr Ala Glu Tyr Thr
               1535                1540                1545
Ser Ile Thr Asp Cys Ile Asp Thr Arg Cys Val Asn Ala Pro Gln
               1550                1555                1560
Ala Ile Ala Asp Arg Ala Ala Phe Pro Gly Gly Leu Gly Asp Lys
               1565                1570                1575
Val Glu Asp Leu Thr Cys Cys His Pro Glu Arg Glu Ala Glu Leu
               1580                1585                1590
Ser His Pro Ser Ser Asp Ser Glu Glu Asn Glu Ala Lys Gly Arg
               1595                1600                1605
Arg Ala Thr Ile Ala Ile Ser Ser Gln Glu Gly Asp Asn Ser Glu
               1610                1615                1620
Arg Thr Leu Ser Asn Asn Ile Thr Val Pro Lys Ile Glu Arg Ala
               1625                1630                1635
Asn Ser Tyr Ser Ala Glu Glu Pro Ser Ala Pro Tyr Ala His Thr
```

```
             1640                1645                1650
Arg Lys Ser Phe Ser Ile Ser Asp Lys Leu Asp Arg Gln Arg Asn
             1655                1660                1665
Thr Ala Ser Leu Arg Asn Pro Phe Gln Arg Ser Lys Ser Ser Lys
             1670                1675                1680
Pro Glu Gly Arg Gly Asp Ser Leu Ser Met Arg Lys Leu Ser Arg
             1685                1690                1695
Thr Ser Ala Phe Gln Ser Phe Glu Ser Lys His Thr
             1700                1705
```

<210> 11
<211> 771
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2493969CD1

<400> 11

```
Met Ser Gly Phe Phe Thr Ser Leu Asp Pro Arg Arg Val Gln Trp
  1               5                  10                  15
Gly Ala Ala Trp Tyr Ala Met His Ser Arg Ile Leu Arg Thr Lys
             20                  25                  30
Pro Val Glu Ser Met Leu Glu Gly Thr Gly Thr Thr Thr Ala His
             35                  40                  45
Gly Thr Lys Leu Ala Gln Val Leu Thr Thr Val Asp Leu Ile Ser
             50                  55                  60
Leu Gly Val Gly Ser Cys Val Gly Thr Gly Met Tyr Val Val Ser
             65                  70                  75
Gly Leu Val Ala Lys Glu Met Ala Gly Pro Gly Val Ile Val Ser
             80                  85                  90
Phe Ile Ile Ala Ala Val Ala Ser Ile Leu Ser Gly Val Cys Tyr
             95                 100                 105
Ala Glu Phe Gly Val Arg Val Pro Lys Thr Thr Gly Ser Ala Tyr
            110                 115                 120
Thr Tyr Ser Tyr Val Thr Val Gly Glu Phe Val Ala Phe Phe Ile
            125                 130                 135
Gly Trp Asn Leu Ile Leu Glu Tyr Leu Ile Gly Thr Ala Ala Gly
            140                 145                 150
Ala Ser Ala Leu Ser Ser Met Phe Asp Ser Leu Ala Asn His Thr
            155                 160                 165
Ile Ser Arg Trp Met Ala Asp Ser Val Gly Thr Leu Asn Gly Leu
            170                 175                 180
Gly Lys Gly Glu Glu Ser Tyr Pro Asp Leu Leu Ala Leu Leu Ile
            185                 190                 195
Ala Val Ile Val Thr Ile Ile Val Ala Leu Gly Val Lys Asn Ser
            200                 205                 210
Ile Gly Phe Asn Asn Val Leu Asn Val Leu Asn Leu Ala Val Trp
            215                 220                 225
Val Phe Ile Met Ile Ala Gly Leu Phe Phe Ile Asn Gly Lys Tyr
            230                 235                 240
Trp Ala Glu Gly Gln Phe Leu Pro His Gly Trp Ser Gly Val Leu
            245                 250                 255
Gln Gly Ala Ala Thr Cys Phe Tyr Ala Phe Ile Gly Phe Asp Ile
            260                 265                 270
```



```
Ile Ala Thr Thr Gly Glu Glu Ala Lys Asn Pro Asn Thr Ser Ile
            275                 280                 285
Pro Tyr Ala Ile Thr Ala Ser Leu Val Ile Cys Leu Thr Ala Tyr
            290                 295                 300
Val Ser Val Ser Val Ile Leu Thr Leu Met Val Pro Tyr Tyr Thr
            305                 310                 315
Ile Asp Thr Glu Ser Pro Leu Met Glu Met Phe Val Ala His Gly
            320                 325                 330
Phe Tyr Ala Ala Lys Phe Val Val Ala Ile Gly Ser Val Ala Gly
            335                 340                 345
Leu Thr Val Ser Leu Leu Gly Ser Leu Phe Pro Met Pro Arg Val
            350                 355                 360
Ile Tyr Ala Met Ala Gly Asp Gly Leu Leu Phe Arg Phe Leu Ala
            365                 370                 375
His Val Ser Ser Tyr Thr Glu Thr Pro Val Val Ala Cys Ile Val
            380                 385                 390
Ser Gly Phe Leu Ala Ala Leu Leu Ala Leu Leu Val Ser Leu Arg
            395                 400                 405
Asp Leu Ile Glu Met Met Ser Ile Gly Thr Leu Leu Ala Tyr Thr
            410                 415                 420
Leu Val Ser Val Cys Val Leu Leu Leu Arg Tyr Gln Pro Glu Ser
            425                 430                 435
Asp Ile Asp Gly Phe Val Lys Phe Leu Ser Glu Glu His Thr Lys
            440                 445                 450
Lys Lys Glu Gly Ile Leu Ala Asp Cys Glu Lys Glu Ala Cys Ser
            455                 460                 465
Pro Val Ser Glu Gly Asp Glu Phe Ser Gly Pro Ala Thr Asn Thr
            470                 475                 480
Cys Gly Ala Lys Asn Leu Pro Ser Leu Gly Asp Asn Glu Met Leu
            485                 490                 495
Ile Gly Lys Ser Asp Lys Ser Thr Tyr Asn Val Asn His Pro Asn
            500                 505                 510
Tyr Gly Thr Val Asp Met Thr Thr Gly Ile Glu Ala Asp Glu Ser
            515                 520                 525
Glu Asn Ile Tyr Leu Ile Lys Leu Lys Lys Leu Ile Gly Pro His
            530                 535                 540
Tyr Tyr Thr Met Arg Ile Arg Leu Gly Leu Pro Gly Lys Met Asp
            545                 550                 555
Arg Pro Thr Ala Ala Thr Gly His Thr Val Thr Ile Cys Val Leu
            560                 565                 570
Leu Leu Phe Ile Leu Met Phe Ile Phe Cys Ser Phe Ile Ile Phe
            575                 580                 585
Gly Ser Asp Tyr Ile Ser Glu Gln Ser Trp Trp Ala Ile Leu Leu
            590                 595                 600
Val Val Leu Met Val Leu Leu Ile Ser Thr Leu Val Phe Val Ile
            605                 610                 615
Leu Gln Gln Pro Glu Asn Pro Lys Lys Leu Pro Tyr Met Ala Pro
            620                 625                 630
Cys Leu Pro Phe Val Pro Ala Phe Ala Met Leu Val Asn Ile Tyr
            635                 640                 645
Leu Met Leu Lys Leu Ser Thr Ile Thr Trp Ile Arg Phe Ala Val
            650                 655                 660
Trp Cys Phe Val Gly Leu Leu Ile Tyr Phe Gly Tyr Gly Ile Trp
            665                 670                 675
Asn Ser Thr Leu Glu Ile Ser Ala Arg Glu Glu Ala Leu His Gln
            680                 685                 690
```

```
Ser Thr Tyr Gln Arg Tyr Asp Val Asp Asp Pro Phe Ser Val Glu
                695                 700                 705
Glu Gly Phe Ser Tyr Ala Thr Glu Gly Glu Ser Gln Glu Asp Trp
                710                 715                 720
Gly Gly Pro Thr Glu Asp Lys Gly Phe Tyr Tyr Gln Gln Met Ser
                725                 730                 735
Asp Ala Lys Ala Asn Gly Arg Thr Ser Ser Lys Ala Lys Ser Lys
                740                 745                 750
Ser Lys His Lys Gln Asn Ser Glu Ala Leu Ile Ala Asn Asp Glu
                755                 760                 765
Leu Asp Tyr Ser Pro Glu
                770
```

<210> 12
<211> 1329
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3244593CD1

<400> 12

```
Met Val Gly Glu Gly Pro Tyr Leu Ile Ser Asp Leu Asp Gln Arg
  1                 5                  10                  15
Gly Arg Arg Arg Ser Phe Ala Glu Arg Tyr Asp Pro Ser Leu Lys
                 20                  25                  30
Thr Met Ile Pro Val Arg Pro Cys Ala Arg Leu Ala Pro Asn Pro
                 35                  40                  45
Val Asp Asp Ala Gly Leu Leu Ser Phe Ala Thr Phe Ser Trp Leu
                 50                  55                  60
Thr Pro Val Met Val Lys Gly Tyr Arg Gln Arg Leu Thr Val Asp
                 65                  70                  75
Thr Leu Pro Pro Leu Ser Thr Tyr Asp Ser Ser Asp Thr Asn Ala
                 80                  85                  90
Lys Arg Phe Arg Val Leu Trp Asp Glu Glu Val Ala Arg Val Gly
                 95                 100                 105
Pro Glu Lys Ala Ser Leu Ser His Val Val Trp Lys Phe Gln Arg
                110                 115                 120
Thr Arg Val Leu Met Asp Ile Val Ala Asn Ile Leu Cys Ile Ile
                125                 130                 135
Met Ala Ala Ile Gly Pro Thr Val Leu Ile His Gln Ile Leu Gln
                140                 145                 150
Gln Thr Glu Arg Thr Ser Gly Lys Val Trp Val Gly Ile Gly Leu
                155                 160                 165
Cys Ile Ala Leu Phe Ala Thr Glu Phe Thr Lys Val Phe Phe Trp
                170                 175                 180
Ala Leu Ala Trp Ala Ile Asn Tyr Arg Thr Ala Ile Arg Leu Lys
                185                 190                 195
Val Ala Leu Ser Thr Leu Val Phe Glu Asn Leu Val Ser Phe Lys
                200                 205                 210
Thr Leu Thr His Ile Ser Val Gly Glu Val Leu Asn Ile Leu Ser
                215                 220                 225
Ser Asp Ser Tyr Ser Leu Phe Glu Ala Ala Leu Phe Cys Pro Leu
                230                 235                 240
Pro Ala Thr Ile Pro Ile Leu Met Val Phe Cys Ala Ala Tyr Ala
```



```
                245                 250                 255
Phe Phe Ile Leu Gly Pro Thr Ala Leu Ile Gly Ile Ser Val Tyr
                260                 265                 270
Val Ile Phe Ile Pro Val Gln Met Phe Met Ala Lys Leu Asn Ser
                275                 280                 285
Ala Phe Arg Arg Ser Ala Ile Leu Val Thr Asp Lys Arg Val Gln
                290                 295                 300
Thr Met Asn Glu Phe Leu Thr Cys Ile Arg Leu Ile Lys Met Tyr
                305                 310                 315
Ala Trp Glu Lys Ser Phe Thr Asn Thr Ile Gln Asp Ile Arg Arg
                320                 325                 330
Arg Glu Arg Lys Leu Leu Glu Lys Ala Gly Phe Val Gln Ser Gly
                335                 340                 345
Asn Ser Ala Leu Ala Pro Ile Val Ser Thr Ile Ala Ile Val Leu
                350                 355                 360
Thr Leu Ser Cys His Ile Leu Leu Arg Arg Lys Leu Thr Ala Pro
                365                 370                 375
Val Ala Phe Ser Val Ile Ala Met Phe Asn Val Met Lys Phe Ser
                380                 385                 390
Ile Ala Ile Leu Pro Phe Ser Ile Lys Ala Met Ala Glu Ala Asn
                395                 400                 405
Val Ser Leu Arg Arg Met Lys Lys Ile Leu Ile Asp Lys Ser Pro
                410                 415                 420
Pro Ser Tyr Ile Thr Gln Pro Glu Asp Pro Asp Thr Val Leu Leu
                425                 430                 435
Leu Ala Asn Ala Thr Leu Thr Trp Glu His Glu Ala Ser Arg Lys
                440                 445                 450
Ser Thr Pro Lys Lys Leu Gln Asn Gln Lys Arg His Leu Cys Lys
                455                 460                 465
Lys Gln Arg Ser Glu Ala Tyr Ser Glu Arg Ser Pro Pro Ala Lys
                470                 475                 480
Gly Ala Thr Gly Pro Glu Glu Gln Ser Asp Ser Leu Lys Ser Val
                485                 490                 495
Leu His Ser Ile Ser Phe Val Val Arg Lys Gly Lys Ile Leu Gly
                500                 505                 510
Ile Cys Gly Asn Val Gly Ser Gly Lys Ser Ser Leu Leu Ala Ala
                515                 520                 525
Leu Leu Gly Gln Met Gln Leu Gln Lys Gly Val Val Ala Val Asn
                530                 535                 540
Gly Thr Leu Ala Tyr Val Ser Gln Gln Ala Trp Ile Phe His Gly
                545                 550                 555
Asn Val Arg Glu Asn Ile Leu Phe Gly Glu Lys Tyr Asp His Gln
                560                 565                 570
Arg Tyr Gln His Thr Val Arg Val Cys Gly Leu Gln Lys Asp Leu
                575                 580                 585
Ser Asn Leu Pro Tyr Gly Asp Leu Thr Glu Ile Gly Glu Arg Gly
                590                 595                 600
Leu Asn Leu Ser Gly Gly Gln Arg Gln Arg Ile Ser Leu Ala Arg
                605                 610                 615
Ala Val Tyr Ser Asp Arg Gln Leu Tyr Leu Leu Asp Asp Pro Leu
                620                 625                 630
Ser Ala Val Asp Ala His Val Gly Lys His Val Phe Glu Glu Cys
                635                 640                 645
Ile Lys Lys Thr Leu Arg Gly Lys Thr Val Val Leu Val Thr His
                650                 655                 660
Gln Leu Gln Phe Leu Glu Ser Cys Asp Glu Val Ile Leu Leu Glu
```

```
                665                 670                 675
Asp Gly Glu Ile Cys Glu Lys Gly Thr His Lys Glu Leu Met Glu
                680                 685                 690
Glu Arg Gly Arg Tyr Ala Lys Leu Ile His Asn Leu Arg Gly Leu
                695                 700                 705
Gln Phe Lys Asp Pro Glu His Leu Tyr Asn Ala Ala Met Val Glu
                710                 715                 720
Ala Phe Lys Glu Ser Pro Ala Glu Arg Glu Glu Asp Ala Gly Ile
                725                 730                 735
Ile Val Leu Ala Pro Gly Asn Glu Lys Asp Glu Gly Lys Glu Ser
                740                 745                 750
Glu Thr Gly Ser Glu Phe Val Asp Thr Lys Gly Tyr Leu Leu Ser
                755                 760                 765
Leu Phe Thr Val Phe Leu Phe Leu Leu Met Ile Gly Ser Ala Ala
                770                 775                 780
Phe Ser Asn Trp Trp Leu Gly Leu Trp Leu Asp Lys Gly Ser Arg
                785                 790                 795
Met Thr Cys Gly Pro Gln Gly Asn Arg Thr Met Cys Glu Val Gly
                800                 805                 810
Ala Val Leu Ala Asp Ile Gly Gln His Val Tyr Gln Trp Val Tyr
                815                 820                 825
Thr Ala Ser Met Val Phe Met Leu Val Phe Gly Val Thr Lys Gly
                830                 835                 840
Phe Val Phe Thr Lys Thr Thr Leu Met Ala Ser Ser Ser Leu His
                845                 850                 855
Asp Thr Val Phe Asp Lys Ile Leu Lys Ser Pro Met Ser Phe Phe
                860                 865                 870
Asp Thr Thr Pro Thr Gly Arg Leu Met Asn Arg Phe Ser Lys Asp
                875                 880                 885
Met Asp Glu Leu Asp Val Arg Leu Pro Phe His Ala Glu Asn Phe
                890                 895                 900
Leu Gln Gln Phe Phe Met Val Val Phe Ile Leu Val Ile Leu Ala
                905                 910                 915
Ala Val Phe Pro Ala Val Leu Leu Val Val Ala Ser Leu Ala Val
                920                 925                 930
Gly Phe Phe Ile Leu Leu Arg Ile Phe His Arg Gly Val Gln Glu
                935                 940                 945
Leu Lys Lys Val Glu Asn Val Ser Arg Ser Pro Trp Phe Thr His
                950                 955                 960
Ile Thr Ser Ser Met Gln Gly Leu Gly Ile Ile His Ala Tyr Gly
                965                 970                 975
Lys Lys Glu Ser Cys Ile Thr Tyr His Leu Leu Tyr Phe Asn Cys
                980                 985                 990
Ala Leu Arg Trp Phe Ala Leu Arg Met Asp Val Leu Met Asn Ile
                995                1000                1005
Leu Thr Phe Thr Val Ala Leu Leu Val Thr Leu Ser Phe Ser Ser
               1010                1015                1020
Ile Ser Thr Ser Ser Lys Gly Leu Ser Leu Ser Tyr Ile Ile Gln
               1025                1030                1035
Leu Ser Gly Leu Leu Gln Val Cys Val Arg Thr Gly Thr Glu Thr
               1040                1045                1050
Gln Ala Lys Phe Thr Ser Val Glu Leu Leu Arg Glu Tyr Ile Ser
               1055                1060                1065
Thr Cys Val Pro Glu Cys Thr His Pro Leu Lys Val Gly Thr Cys
               1070                1075                1080
Pro Lys Asp Trp Pro Ser Cys Gly Glu Ile Thr Phe Arg Asp Tyr
```



```
               1085                1090                1095
Gln Met Arg Tyr Arg Asp Asn Thr Pro Leu Val Leu Asp Ser Leu
               1100                1105                1110
Asn Leu Asn Ile Gln Ser Gly Gln Thr Val Gly Ile Val Gly Arg
               1115                1120                1125
Thr Gly Ser Gly Lys Ser Ser Leu Gly Met Ala Leu Phe Arg Leu
               1130                1135                1140
Val Glu Pro Ala Ser Gly Thr Ile Phe Ile Asp Glu Val Asp Ile
               1145                1150                1155
Cys Ile Leu Ser Leu Glu Asp Leu Arg Thr Lys Leu Thr Val Ile
               1160                1165                1170
Pro Gln Asp Pro Val Leu Phe Val Gly Thr Val Arg Tyr Asn Leu
               1175                1180                1185
Asp Pro Phe Glu Ser His Thr Asp Glu Met Leu Trp Gln Val Leu
               1190                1195                1200
Glu Arg Thr Phe Met Arg Asp Thr Ile Met Lys Leu Pro Glu Lys
               1205                1210                1215
Leu Gln Ala Glu Val Thr Glu Asn Gly Glu Asn Phe Ser Val Gly
               1220                1225                1230
Glu Arg Gln Leu Leu Cys Val Ala Arg Ala Leu Leu Arg Asn Ser
               1235                1240                1245
Lys Ile Ile Leu Leu Asp Glu Ala Thr Ala Ser Met Asp Ser Lys
               1250                1255                1260
Thr Asp Thr Leu Val Gln Asn Thr Ile Lys Asp Ala Phe Lys Gly
               1265                1270                1275
Cys Thr Val Leu Thr Ile Ala His Arg Leu Asn Thr Val Leu Asn
               1280                1285                1290
Cys Asp His Val Leu Val Met Glu Asn Gly Lys Val Ile Glu Phe
               1295                1300                1305
Asp Lys Pro Glu Val Leu Ala Glu Lys Pro Asp Ser Ala Phe Ala
               1310                1315                1320
Met Leu Leu Ala Ala Glu Val Arg Leu
               1325
```

<210> 13
<211> 1353
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4921451CD1

<400> 13

```
Met Gly Thr Gly Pro Ala Gln Thr Pro Arg Ser Thr Arg Ala Gly
  1               5                  10                  15
Pro Glu Pro Ser Pro Ala Pro Pro Gly Pro Gly Asp Thr Gly Asp
             20                  25                  30
Ser Asp Val Thr Gln Glu Gly Ser Gly Pro Ala Gly Ile Arg Gly
             35                  40                  45
Ala Pro Pro Ala Trp Ala Ala Ser Ala Arg Glu Lys Ile Ser Glu
             50                  55                  60
Met Arg Thr Gly Thr Gln Val Leu Ile Leu Gly Gly Gly Gly Gly
             65                  70                  75
Ala Ala Phe Thr Trp Lys Val Gln Ala Asn Asn Arg Ala Tyr Asn
             80                  85                  90
```

```
Gly Gln Phe Lys Glu Lys Val Ile Leu Cys Trp Gln Arg Lys Lys
                95                  100                 105
Tyr Lys Thr Asn Val Ile Arg Thr Ala Lys Tyr Asn Phe Tyr Ser
            110                 115                 120
Phe Leu Pro Leu Asn Leu Tyr Glu Gln Phe His Arg Val Ser Asn
        125                 130                 135
Leu Phe Phe Leu Ile Ile Ile Ile Leu Gln Ser Ile Pro Asp Ile
    140                 145                 150
Ser Thr Leu Pro Trp Phe Ser Leu Ser Thr Pro Met Val Cys Leu
155                 160                 165
Leu Phe Ile Arg Ala Thr Arg Asp Leu Val Asp Asp Met Gly Arg
            170                 175                 180
His Lys Ser Asp Arg Ala Ile Asn Asn Arg Pro Cys Gln Ile Leu
        185                 190                 195
Met Gly Lys Ser Phe Lys Gln Lys Lys Trp Gln Asp Leu Cys Val
    200                 205                 210
Gly Asp Val Val Cys Leu Arg Lys Asp Asn Ile Val Pro Val Ser
215                 220                 225
Trp Gly Gly Pro Arg Gly Pro Arg Thr Thr Arg Pro Leu Thr Glu
            230                 235                 240
Ser Thr Pro Pro Arg Val Gly Arg Ala Ala Ala Pro Pro Ile Cys
        245                 250                 255
Leu Ala Ser Pro Leu Ala Thr Leu Pro Pro Thr Pro His Gln Ala
    260                 265                 270
Asp Met Leu Leu Leu Ala Ser Thr Glu Pro Ser Ser Leu Cys Tyr
275                 280                 285
Val Glu Thr Val Asp Ile Asp Gly Glu Thr Asn Leu Lys Phe Arg
            290                 295                 300
Gln Ala Leu Met Val Thr His Lys Glu Leu Ala Thr Ile Lys Lys
        305                 310                 315
Met Ala Ser Phe Gln Gly Thr Val Thr Cys Glu Ala Pro Asn Ser
    320                 325                 330
Arg Met His His Phe Val Gly Cys Leu Glu Trp Asn Asp Lys Lys
335                 340                 345
Tyr Ser Leu Asp Ile Gly Asn Leu Leu Leu Arg Gly Cys Arg Ile
            350                 355                 360
Arg Asn Thr Asp Thr Cys Tyr Gly Leu Val Ile Tyr Ala Gly Phe
        365                 370                 375
Asp Thr Lys Ile Met Lys Asn Cys Gly Lys Ile His Leu Lys Arg
    380                 385                 390
Thr Lys Leu Asp Leu Leu Met Asn Lys Leu Val Val Val Ile Phe
395                 400                 405
Ile Ser Val Val Leu Val Cys Leu Val Leu Ala Phe Gly Phe Gly
            410                 415                 420
Phe Ser Val Lys Glu Phe Lys Asp His His Tyr Tyr Leu Ser Gly
        425                 430                 435
Val His Gly Ser Ser Val Ala Ala Glu Ser Phe Phe Val Phe Trp
    440                 445                 450
Ser Phe Leu Ile Leu Leu Ser Val Thr Ile Pro Met Ser Met Phe
455                 460                 465
Ile Leu Ser Glu Phe Ile Tyr Leu Gly Asn Ser Val Phe Ile Asp
            470                 475                 480
Trp Asp Val Gln Met Tyr Tyr Lys Pro Gln Asp Val Pro Ala Lys
        485                 490                 495
Ala Arg Ser Thr Ser Leu Asn Asp His Leu Gly Gln Val Glu Tyr
    500                 505                 510
```



```
Ile Phe Ser Asp Lys Thr Gly Thr Leu Thr Gln Asn Ile Leu Thr
                515                 520                 525
Phe Asn Lys Cys Cys Ile Ser Gly Arg Val Tyr Gly Glu Pro Leu
            530                 535                 540
Pro Leu Glu Gln Val Arg Arg Arg Glu Ala Ala Leu Pro Gln Cys
        545                 550                 555
Gly Pro Ala Ala Pro Arg Ala Asp Gln Arg Gly Arg Gly Arg Ala
    560                 565                 570
Gly Val Leu Ala Pro Ala Gly His Leu Pro His Gly Asp Asp Gln
575                 580                 585
Leu Leu Tyr Gln Ala Ala Ser Pro Asp Glu Gly Ala Leu Val Thr
            590                 595                 600
Ala Ala Arg Asn Phe Gly Tyr Val Phe Leu Ser Arg Thr Gln Asp
        605                 610                 615
Thr Val Thr Ile Met Glu Leu Gly Glu Glu Arg Val Tyr Gln Val
    620                 625                 630
Leu Ala Ile Met Asp Phe Asn Ser Thr Arg Lys Arg Met Ser Val
635                 640                 645
Leu Val Arg Lys Pro Glu Gly Ala Ile Cys Leu Tyr Thr Lys Gly
            650                 655                 660
Ala Asp Thr Val Ile Phe Glu Arg Leu His Arg Arg Gly Ala Met
        665                 670                 675
Glu Phe Ala Thr Glu Glu Ala Leu Ala Ala Phe Ala Gln Glu Thr
    680                 685                 690
Leu Arg Thr Leu Cys Leu Ala Tyr Arg Glu Val Ala Glu Asp Ile
695                 700                 705
Tyr Glu Asp Trp Gln Gln Arg His Gln Glu Ala Ser Leu Leu Leu
            710                 715                 720
Gln Asn Arg Ala Gln Ala Leu Gln Gln Val Tyr Asn Glu Met Glu
        725                 730                 735
Gln Asp Leu Arg Leu Leu Gly Ala Thr Ala Ile Glu Asp Arg Leu
    740                 745                 750
Gln Asp Gly Val Pro Glu Thr Ile Lys Cys Leu Lys Lys Ser Asn
755                 760                 765
Ile Lys Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr Ala Val
            770                 775                 780
Asn Ile Gly Phe Ala Cys Glu Leu Leu Ser Glu Asn Met Leu Ile
        785                 790                 795
Leu Glu Glu Lys Glu Ile Ser Arg Ile Leu Glu Thr Tyr Trp Glu
    800                 805                 810
Asn Ser Asn Asn Leu Leu Thr Arg Glu Ser Leu Ser Gln Val Lys
815                 820                 825
Leu Ala Leu Val Ile Asn Gly Asp Phe Leu Asp Lys Leu Leu Val
            830                 835                 840
Ser Leu Arg Lys Glu Pro Arg Ala Leu Ala Gln Asn Val Asn Met
        845                 850                 855
Asp Glu Ala Trp Gln Glu Leu Gly Gln Ser Arg Arg Asp Phe Leu
    860                 865                 870
Tyr Ala Arg Arg Leu Ser Leu Leu Cys Arg Arg Phe Gly Leu Pro
875                 880                 885
Leu Ala Ala Pro Pro Ala Gln Asp Ser Arg Ala Arg Arg Ser Ser
            890                 895                 900
Glu Val Leu Gln Glu Arg Ala Phe Val Asp Leu Ala Ser Lys Cys
        905                 910                 915
Gln Ala Val Ile Cys Cys Arg Val Thr Pro Lys Gln Lys Ala Leu
    920                 925                 930
```

Ile Val Ala Leu Val Lys Lys Tyr His Gln Val Val Thr Leu Ala
935 940 945
Ile Gly Asp Gly Ala Asn Asp Ile Asn Met Ile Lys Thr Ala Asp
950 955 960
Val Gly Val Gly Leu Ala Gly Gln Glu Gly Met Gln Ala Val Gln
965 970 975
Asn Ser Asp Phe Val Leu Gly Gln Phe Cys Phe Leu Gln Arg Leu
980 985 990
Leu Leu Val His Gly Arg Trp Ser Tyr Val Arg Ile Cys Lys Phe
995 1000 1005
Leu Arg Tyr Phe Phe Tyr Lys Ser Met Ala Ser Met Met Val Gln
1010 1015 1020
Val Trp Phe Ala Cys Tyr Asn Gly Phe Thr Gly Gln Asp Val Ser
1025 1030 1035
Ala Glu Gln Ser Leu Glu Lys Pro Glu Leu Tyr Val Val Gly Gln
1040 1045 1050
Lys Asp Glu Leu Phe Asn Tyr Trp Val Phe Val Gln Ala Ile Ala
1055 1060 1065
His Gly Val Thr Thr Ser Leu Val Asn Phe Phe Met Thr Leu Trp
1070 1075 1080
Ile Ser Arg Asp Thr Ala Gly Pro Ala Ser Phe Ser Asp His Gln
1085 1090 1095
Ser Phe Ala Val Val Val Ala Leu Ser Cys Leu Leu Ser Ile Thr
1100 1105 1110
Met Glu Val Ile Leu Ile Ile Lys Tyr Trp Thr Ala Leu Cys Val
1115 1120 1125
Ala Thr Ile Leu Leu Ser Leu Gly Phe Tyr Ala Ile Met Thr Thr
1130 1135 1140
Thr Thr Gln Ser Phe Trp Leu Phe Arg Val Ser Pro Thr Thr Phe
1145 1150 1155
Pro Phe Leu Tyr Ala Asp Leu Ser Val Met Ser Ser Pro Ser Ile
1160 1165 1170
Leu Leu Val Val Leu Leu Ser Val Ser Ile Asn Thr Phe Pro Val
1175 1180 1185
Leu Ala Leu Arg Val Ile Phe Pro Ala Leu Lys Glu Leu Arg Ala
1190 1195 1200
Lys Glu Glu Lys Val Glu Glu Gly Pro Ser Glu Glu Ile Phe Thr
1205 1210 1215
Met Glu Pro Leu Pro His Val His Arg Glu Ser Arg Ala Arg Arg
1220 1225 1230
Ser Ser Tyr Ala Phe Ser His Arg Gln Leu Thr Leu Glu Ser Gln
1235 1240 1245
Pro Asp Ser Ser Glu Glu Lys Ser Ala Phe Leu Lys Pro Ser Thr
1250 1255 1260
Pro Phe Arg Lys Ser Trp Gln Lys Glu Pro His Thr Pro Lys Glu
1265 1270 1275
Gly Thr Val Pro Leu Pro Asp Lys Thr His Lys Ser Gln Val Glu
1280 1285 1290
Thr Leu Pro Pro Ser Leu Glu Glu Ser Ser Thr Ser Thr Ser Glu
1295 1300 1305
Gln Pro Met Glu Val Glu Leu Trp Pro Ala Glu Lys Gln Ser Ser
1310 1315 1320
Ser Ser Met Glu Trp Leu Leu Val Pro Gly Glu Glu Gln Leu Ser
1325 1330 1335
Leu Pro Pro Glu Glu Gln Ser Leu Pro Ser Ala Glu Gly Thr Arg
1340 1345 1350

Val Gln Gln

<210> 14
<211> 921
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5547443CD1

<400> 14
Met Ala His Glu Ser Ala Glu Asp Leu Phe His Phe Asn Val Gly
1 5 10 15
Gly Trp His Phe Ser Val Pro Arg Ser Lys Leu Ser Gln Phe Pro
20 25 30
Asp Ser Leu Leu Trp Lys Glu Ala Ser Ala Leu Thr Ser Ser Glu
35 40 45
Ser Gln Arg Leu Phe Ile Asp Arg Asp Gly Ser Thr Phe Arg His
50 55 60
Val His Tyr Tyr Leu Tyr Thr Ser Lys Leu Ser Phe Ser Ser Cys
65 70 75
Ala Glu Leu Asn Leu Leu Tyr Glu Gln Ala Leu Gly Leu Gln Leu
80 85 90
Met Pro Leu Leu Gln Thr Leu Asp Asn Leu Lys Glu Gly Lys His
95 100 105
His Leu Arg Val Arg Pro Ala Asp Leu Pro Val Ala Glu Arg Ala
110 115 120
Ser Leu Asn Tyr Trp Arg Thr Trp Lys Cys Ile Ser Lys Pro Ser
125 130 135
Glu Phe Pro Ile Lys Ser Pro Ala Phe Thr Gly Leu His Asp Lys
140 145 150
Ala Pro Leu Gly Leu Met Asp Thr Pro Leu Leu Asp Thr Glu Glu
155 160 165
Glu Val His Tyr Cys Phe Leu Pro Leu Asp Leu Val Ala Lys Tyr
170 175 180
Pro Ser Leu Val Thr Glu Asp Asn Leu Leu Trp Leu Ala Glu Thr
185 190 195
Val Ala Leu Ile Glu Cys Glu Cys Ser Glu Phe Arg Phe Ile Val
200 205 210
Asn Phe Leu Arg Ser Gln Lys Ile Leu Leu Pro Asp Asn Phe Ser
215 220 225
Asn Ile Asp Val Leu Glu Ala Glu Val Glu Ile Leu Glu Ile Pro
230 235 240
Ala Leu Thr Glu Ala Val Arg Trp Tyr Arg Met Asn Met Gly Gly
245 250 255
Cys Ser Pro Thr Thr Cys Ser Pro Leu Ser Pro Gly Lys Gly Ala
260 265 270
Arg Thr Ala Ser Leu Glu Ser Val Lys Pro Leu Tyr Thr Met Ala
275 280 285
Leu Gly Leu Leu Val Lys Tyr Pro Asp Ser Ala Leu Gly Gln Leu
290 295 300
Arg Ile Glu Ser Thr Leu Asp Gly Ser Arg Leu Tyr Ile Thr Gly
305 310 315
Asn Gly Val Leu Phe Gln His Val Lys Asn Trp Leu Gly Thr Cys

```
                320                 325                 330
Arg Leu Pro Leu Thr Glu Thr Ile Ser Glu Val Tyr Glu Leu Cys
                335                 340                 345
Ala Phe Leu Asp Lys Arg Asp Ile Thr Tyr Glu Pro Ile Lys Val
                350                 355                 360
Ala Leu Lys Thr His Leu Glu Pro Arg Thr Leu Ala Pro Met Asp
                365                 370                 375
Val Leu Asn Glu Trp Thr Ala Glu Ile Thr Val Tyr Ser Pro Gln
                380                 385                 390
Gln Ile Ile Lys Val Tyr Val Gly Ser His Trp Tyr Ala Thr Thr
                395                 400                 405
Leu Gln Thr Leu Leu Lys Tyr Pro Glu Leu Leu Ser Asn Pro Gln
                410                 415                 420
Arg Val Tyr Trp Ile Thr Tyr Gly Gln Thr Leu Leu Ile His Gly
                425                 430                 435
Asp Gly Gln Met Phe Arg His Ile Leu Asn Phe Leu Arg Leu Gly
                440                 445                 450
Lys Leu Phe Leu Pro Ser Glu Phe Lys Glu Trp Pro Leu Phe Cys
                455                 460                 465
Gln Glu Val Glu Glu Tyr His Ile Pro Ser Leu Ser Glu Ala Leu
                470                 475                 480
Ala Gln Cys Glu Ala Tyr Lys Ser Trp Thr Gln Glu Lys Glu Ser
                485                 490                 495
Glu Asn Glu Glu Ala Phe Ser Ile Arg Arg Leu His Val Val Thr
                500                 505                 510
Glu Gly Pro Gly Ser Leu Val Glu Phe Ser Arg Asp Thr Lys Glu
                515                 520                 525
Thr Thr Ala Tyr Met Pro Val Asp Phe Glu Asp Cys Ser Asp Arg
                530                 535                 540
Thr Pro Trp Asn Lys Ala Lys Gly Asn Leu Val Arg Ser Asn Gln
                545                 550                 555
Met Asp Glu Ala Glu Gln Tyr Thr Arg Pro Ile Gln Val Ser Leu
                560                 565                 570
Cys Arg Asn Ala Lys Arg Ala Gly Asn Pro Ser Thr Tyr Ser His
                575                 580                 585
Cys Arg Gly Leu Cys Thr Asn Pro Gly His Trp Gly Ser His Pro
                590                 595                 600
Glu Ser Pro Pro Lys Lys Lys Cys Thr Thr Ile Asn Leu Thr Gln
                605                 610                 615
Lys Ser Glu Thr Lys Asp Pro Pro Ala Thr Pro Met Gln Lys Leu
                620                 625                 630
Ile Ser Leu Val Arg Glu Trp Asp Met Val Asn Cys Lys Gln Trp
                635                 640                 645
Glu Phe Gln Pro Leu Thr Ala Thr Arg Ser Ser Pro Leu Glu Glu
                650                 655                 660
Ala Thr Leu Gln Leu Pro Leu Gly Ser Glu Ala Ala Ser Gln Pro
                665                 670                 675
Ser Thr Ser Ala Ala Trp Lys Ala His Ser Thr Ala Ser Glu Lys
                680                 685                 690
Asp Pro Gly Pro Gln Ala Gly Ala Gly Ala Gly Ala Lys Asp Lys
                695                 700                 705
Gly Pro Glu Pro Thr Phe Lys Pro Tyr Leu Pro Pro Lys Arg Ala
                710                 715                 720
Gly Thr Leu Lys Asp Trp Ser Lys Gln Arg Thr Lys Glu Arg Glu
                725                 730                 735
Ser Pro Ala Pro Glu Gln Pro Leu Pro Glu Ala Ser Glu Val Asp
```



```
                740                 745                 750
Ser Leu Gly Val Ile Leu Lys Val Thr His Pro Pro Val Val Gly
                755                 760                 765
Ser Asp Gly Phe Cys Met Phe Phe Glu Asp Ser Ile Ile Tyr Thr
                770                 775                 780
Thr Glu Met Asp Asn Leu Arg His Thr Thr Pro Thr Ala Ser Pro
                785                 790                 795
Gln Pro Gln Glu Val Thr Phe Leu Ser Phe Ser Leu Ser Trp Glu
                800                 805                 810
Glu Met Phe Tyr Ala Gln Lys Cys His Cys Phe Leu Ala Asp Ile
                815                 820                 825
Ile Met Asp Ser Ile Arg Gln Lys Asp Pro Lys Ala Ile Thr Ala
                830                 835                 840
Lys Val Val Ser Leu Ala Asn Arg Leu Trp Thr Leu His Ile Ser
                845                 850                 855
Pro Lys Gln Phe Val Val Asp Leu Leu Ala Ile Thr Gly Phe Lys
                860                 865                 870
Asp Asp Arg His Thr Gln Glu Arg Leu Tyr Ser Trp Val Glu Leu
                875                 880                 885
Thr Leu Pro Phe Ala Arg Lys Tyr Gly Arg Cys Met Asp Leu Leu
                890                 895                 900
Ile Gln Arg Gly Leu Ser Arg Ser Val Ser Tyr Ser Ile Leu Gly
                905                 910                 915
Lys Tyr Leu Gln Glu Asp
                920
```

<210> 15
<211> 530
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 56008413CD1

<400> 15

```
Met Gly Ser Val Gly Ser Gln Arg Leu Glu Glu Pro Ser Val Ala
  1               5                  10                  15
Gly Thr Pro Asp Pro Gly Val Val Met Ser Phe Thr Phe Asp Ser
                 20                  25                  30
His Gln Leu Glu Glu Ala Ala Glu Ala Ala Gln Gly Gln Gly Leu
                 35                  40                  45
Arg Ala Arg Gly Val Pro Ala Phe Thr Asp Thr Thr Leu Asp Glu
                 50                  55                  60
Pro Val Pro Asp Asp Arg Tyr His Ala Ile Tyr Phe Ala Met Leu
                 65                  70                  75
Leu Ala Gly Val Gly Phe Leu Leu Pro Tyr Asn Ser Phe Ile Thr
                 80                  85                  90
Asp Val Asp Tyr Leu His His Lys Tyr Pro Gly Thr Ser Ile Val
                 95                 100                 105
Phe Asp Met Ser Leu Thr Tyr Ile Leu Val Ala Leu Ala Ala Val
                110                 115                 120
Leu Leu Asn Asn Val Leu Val Glu Arg Leu Thr Leu His Thr Arg
                125                 130                 135
Ile Thr Ala Gly Tyr Leu Leu Ala Leu Gly Pro Leu Leu Phe Ile
                140                 145                 150
```

```
Ser Ile Cys Asp Val Trp Leu Gln Leu Phe Ser Arg Asp Gln Ala
                155                 160                 165
Tyr Ala Ile Asn Leu Ala Ala Val Gly Thr Val Ala Phe Gly Cys
                170                 175                 180
Thr Val Gln Gln Ser Ser Phe Tyr Gly Tyr Thr Gly Met Leu Pro
                185                 190                 195
Lys Arg Tyr Thr Gln Gly Val Met Thr Gly Glu Ser Thr Ala Gly
                200                 205                 210
Val Met Ile Ser Leu Ser Arg Ile Leu Thr Lys Leu Leu Leu Pro
                215                 220                 225
Asp Glu Arg Ala Ser Thr Leu Ile Phe Phe Leu Val Ser Val Ala
                230                 235                 240
Leu Glu Leu Leu Cys Phe Leu Leu His Leu Leu Val Arg Arg Ser
                245                 250                 255
Arg Phe Val Leu Phe Tyr Thr Thr Arg Pro Arg Asp Ser His Arg
                260                 265                 270
Gly Arg Pro Gly Leu Gly Arg Gly Tyr Gly Tyr Arg Val His His
                275                 280                 285
Asp Val Val Ala Gly Asp Val His Phe Glu His Pro Ala Pro Ala
                290                 295                 300
Leu Ala Pro Asn Glu Ser Pro Lys Asp Ser Pro Ala His Glu Val
                305                 310                 315
Thr Gly Ser Gly Gly Ala Tyr Met Arg Phe Asp Val Pro Arg Pro
                320                 325                 330
Arg Val Gln Arg Ser Trp Pro Thr Phe Arg Ala Leu Leu Leu His
                335                 340                 345
Arg Tyr Val Val Ala Arg Val Ile Trp Ala Asp Met Leu Ser Ile
                350                 355                 360
Ala Val Thr Tyr Phe Ile Thr Leu Cys Leu Phe Pro Gly Leu Glu
                365                 370                 375
Ser Glu Ile Arg His Cys Ile Leu Gly Glu Trp Leu Pro Ile Leu
                380                 385                 390
Ile Met Ala Val Phe Asn Leu Ser Asp Phe Val Gly Lys Ile Leu
                395                 400                 405
Ala Ala Leu Pro Val Asp Trp Arg Gly Thr His Leu Leu Ala Cys
                410                 415                 420
Ser Cys Leu Arg Val Val Phe Ile Pro Leu Phe Ile Leu Cys Val
                425                 430                 435
Tyr Pro Ser Gly Met Pro Ala Leu Arg His Pro Ala Trp Pro Cys
                440                 445                 450
Ile Phe Ser Leu Leu Met Gly Ile Ser Asn Gly Tyr Phe Gly Ser
                455                 460                 465
Val Pro Met Ile Leu Ala Ala Gly Lys Val Ser Pro Lys Gln Arg
                470                 475                 480
Glu Leu Ala Gly Asn Thr Met Thr Val Ser Tyr Met Ser Gly Leu
                485                 490                 495
Thr Leu Gly Ser Ala Val Ala Tyr Cys Thr Tyr Ser Leu Thr Arg
                500                 505                 510
Asp Ala His Gly Ser Cys Leu His Ala Ser Thr Ala Asn Gly Ser
                515                 520                 525
Ile Leu Ala Gly Leu
                530
```

<210> 16
<211> 1617
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6127911CD1

<400> 16

```
Met Asn Met Lys Gln Lys Ser Val Tyr Gln Gln Thr Lys Ala Leu
  1               5                  10                  15
Leu Cys Lys Asn Phe Leu Lys Lys Trp Arg Met Lys Arg Glu Ser
                20                  25                  30
Leu Leu Glu Trp Gly Leu Ser Ile Leu Leu Gly Leu Cys Ile Ala
                35                  40                  45
Leu Phe Ser Ser Ser Met Arg Asn Val Gln Phe Pro Gly Met Ala
                50                  55                  60
Pro Gln Asn Leu Gly Arg Val Asp Lys Phe Asn Ser Ser Ser Leu
                65                  70                  75
Met Val Val Tyr Thr Pro Ile Ser Asn Leu Thr Gln Gln Ile Met
                80                  85                  90
Asn Lys Thr Ala Leu Ala Pro Leu Leu Lys Gly Thr Ser Val Ile
                95                 100                 105
Gly Ala Pro Asn Lys Thr His Met Asp Glu Ile Leu Leu Glu Asn
               110                 115                 120
Leu Pro Tyr Ala Met Gly Ile Ile Phe Asn Glu Thr Phe Ser Tyr
               125                 130                 135
Lys Leu Ile Phe Phe Gln Gly Tyr Asn Ser Pro Leu Trp Lys Glu
               140                 145                 150
Asp Phe Ser Ala His Cys Trp Asp Gly Tyr Gly Glu Phe Ser Cys
               155                 160                 165
Thr Leu Thr Lys Tyr Trp Asn Arg Gly Phe Val Ala Leu Gln Thr
               170                 175                 180
Ala Ile Asn Thr Ala Ile Ile Glu Ile Thr Thr Asn His Pro Val
               185                 190                 195
Met Glu Glu Leu Met Ser Val Thr Ala Ile Thr Met Lys Thr Leu
               200                 205                 210
Pro Phe Ile Thr Lys Asn Leu Leu His Asn Glu Met Phe Ile Leu
               215                 220                 225
Phe Phe Leu Leu His Phe Ser Pro Leu Val Tyr Phe Ile Ser Leu
               230                 235                 240
Asn Val Thr Lys Glu Arg Lys Lys Ser Lys Asn Leu Met Lys Met
               245                 250                 255
Met Gly Leu Gln Asp Ser Ala Phe Trp Leu Ser Trp Gly Leu Ile
               260                 265                 270
Tyr Ala Gly Phe Ile Phe Ile Ile Ser Ile Phe Ile Thr Ile Ile
               275                 280                 285
Ile Thr Phe Thr Gln Ile Ile Val Met Thr Gly Phe Met Val Ile
               290                 295                 300
Phe Ile Leu Phe Phe Leu Tyr Gly Leu Ser Leu Val Ala Leu Val
               305                 310                 315
Phe Leu Met Ser Val Leu Leu Lys Lys Ala Val Leu Thr Asn Leu
               320                 325                 330
Val Val Phe Leu Leu Thr Leu Phe Trp Gly Cys Leu Gly Phe Thr
               335                 340                 345
Val Phe Tyr Glu Gln Leu Pro Ser Ser Leu Glu Trp Ile Leu Asn
               350                 355                 360
Ile Cys Ser Pro Phe Ala Phe Thr Thr Gly Met Ile Gln Ile Ile
```

```
                365                 370                 375
Lys Leu Asp Tyr Asn Leu Asn Gly Val Ile Phe Pro Asp Pro Ser
                380                 385                 390
Gly Asp Ser Tyr Thr Met Ile Ala Thr Phe Ser Met Leu Leu Leu
                395                 400                 405
Asp Gly Leu Ile Tyr Leu Leu Leu Ala Leu Tyr Phe Asp Lys Ile
                410                 415                 420
Leu Pro Tyr Gly Asp Glu Arg His Tyr Ser Pro Leu Phe Phe Leu
                425                 430                 435
Asn Ser Ser Ser Cys Phe Gln His Gln Arg Thr Asn Ala Lys Val
                440                 445                 450
Ile Glu Lys Glu Ile Asp Ala Glu His Pro Ser Asp Asp Tyr Phe
                455                 460                 465
Glu Pro Val Ala Pro Glu Phe Gln Gly Lys Glu Ala Ile Arg Ile
                470                 475                 480
Arg Asn Val Lys Lys Glu Tyr Lys Gly Lys Ser Gly Lys Val Glu
                485                 490                 495
Ala Leu Lys Gly Leu Leu Phe Asp Ile Tyr Glu Gly Gln Ile Thr
                500                 505                 510
Ala Ile Leu Gly His Ser Gly Ala Gly Lys Ser Ser Leu Leu Asn
                515                 520                 525
Ile Leu Asn Gly Leu Ser Val Pro Thr Glu Gly Ser Val Thr Ile
                530                 535                 540
Tyr Asn Lys Asn Leu Ser Glu Met Gln Asp Leu Glu Glu Ile Arg
                545                 550                 555
Lys Ile Thr Gly Val Cys Pro Gln Phe Asn Val Gln Phe Asp Ile
                560                 565                 570
Leu Thr Val Lys Glu Asn Leu Ser Leu Phe Ala Lys Ile Lys Gly
                575                 580                 585
Ile His Leu Lys Glu Val Glu Gln Glu Val Gln Arg Ile Leu Leu
                590                 595                 600
Glu Leu Asp Met Gln Asn Ile Gln Asp Asn Leu Ala Lys His Leu
                605                 610                 615
Ser Glu Gly Gln Lys Arg Lys Leu Thr Phe Gly Ile Thr Ile Leu
                620                 625                 630
Gly Asp Pro Gln Ile Leu Leu Leu Asp Glu Pro Thr Thr Gly Leu
                635                 640                 645
Asp Pro Phe Ser Arg Asp Gln Val Trp Ser Leu Leu Arg Glu Arg
                650                 655                 660
Arg Ala Asp His Val Ile Leu Phe Ser Thr Gln Ser Met Asp Glu
                665                 670                 675
Ala Asp Ile Leu Ala Asp Arg Lys Val Ile Met Ser Asn Gly Arg
                680                 685                 690
Leu Lys Cys Ala Gly Ser Ser Met Phe Leu Lys Arg Arg Trp Gly
                695                 700                 705
Leu Gly Tyr His Leu Ser Leu His Arg Asn Glu Ile Cys Asn Pro
                710                 715                 720
Glu Gln Ile Thr Ser Phe Ile Thr His His Ile Pro Asp Ala Lys
                725                 730                 735
Leu Lys Thr Glu Asn Lys Glu Lys Leu Val Tyr Thr Leu Pro Leu
                740                 745                 750
Glu Arg Thr Asn Thr Phe Pro Asp Leu Phe Ser Asp Leu Asp Lys
                755                 760                 765
Cys Ser Asp Gln Gly Val Thr Gly Tyr Asp Ile Ser Met Ser Thr
                770                 775                 780
Leu Asn Glu Val Phe Met Lys Leu Glu Gly Gln Ser Thr Ile Glu
```



```
                785                 790                 795
Gln Asp Phe Glu Gln Val Glu Met Ile Arg Asp Ser Glu Ser Leu
                800                 805                 810
Asn Glu Met Glu Leu Ala His Ser Ser Phe Ser Glu Met Gln Thr
                815                 820                 825
Ala Val Ser Asp Met Gly Leu Trp Arg Met Gln Val Phe Ala Met
                830                 835                 840
Ala Arg Leu Arg Phe Leu Lys Leu Lys Arg Gln Thr Lys Val Leu
                845                 850                 855
Leu Thr Leu Leu Leu Val Phe Gly Ile Ala Ile Phe Pro Leu Ile
                860                 865                 870
Val Glu Asn Ile Ile Tyr Ala Met Leu Asn Glu Lys Ile Asp Trp
                875                 880                 885
Glu Phe Lys Asn Glu Leu Tyr Phe Leu Ser Pro Gly Gln Leu Pro
                890                 895                 900
Gln Glu Pro Arg Thr Ser Leu Leu Ile Ile Asn Asn Thr Glu Ser
                905                 910                 915
Asn Ile Glu Asp Phe Ile Lys Ser Leu Lys His Gln Asn Ile Leu
                920                 925                 930
Leu Glu Val Asp Asp Phe Glu Asn Arg Asn Gly Thr Asp Gly Leu
                935                 940                 945
Ser Tyr Asn Gly Ala Ile Ile Val Ser Gly Lys Gln Lys Asp Tyr
                950                 955                 960
Arg Phe Ser Val Val Cys Asn Thr Lys Arg Leu His Cys Phe Pro
                965                 970                 975
Ile Leu Met Asn Ile Ile Ser Asn Gly Leu Leu Gln Met Phe Asn
                980                 985                 990
His Thr Gln His Ile Arg Ile Glu Ser Ser Pro Phe Pro Leu Ser
                995                1000                1005
His Ile Gly Leu Trp Thr Gly Leu Pro Asp Gly Ser Phe Phe Leu
               1010                1015                1020
Phe Leu Val Leu Cys Ser Ile Ser Pro Tyr Ile Thr Met Gly Ser
               1025                1030                1035
Ile Ser Asp Tyr Lys Lys Asn Ala Lys Ser Gln Leu Trp Ile Ser
               1040                1045                1050
Gly Leu Tyr Thr Ser Ala Tyr Trp Cys Gly Gln Ala Leu Val Asp
               1055                1060                1065
Val Ser Phe Phe Ile Leu Ile Leu Leu Leu Met Tyr Leu Ile Phe
               1070                1075                1080
Tyr Ile Glu Asn Met Gln Tyr Leu Leu Ile Thr Ser Gln Ile Val
               1085                1090                1095
Phe Ala Leu Val Ile Val Thr Pro Gly Tyr Ala Ala Ser Leu Val
               1100                1105                1110
Phe Phe Ile Tyr Met Ile Ser Phe Ile Phe Arg Lys Arg Arg Lys
               1115                1120                1125
Asn Ser Gly Leu Trp Ser Phe Tyr Phe Phe Phe Ala Ser Thr Ile
               1130                1135                1140
Met Phe Ser Ile Thr Leu Ile Asn His Phe Asp Leu Ser Ile Leu
               1145                1150                1155
Ile Thr Thr Met Val Leu Val Pro Ser Tyr Thr Leu Leu Gly Phe
               1160                1165                1170
Lys Thr Phe Leu Glu Val Arg Asp Gln Glu His Tyr Arg Glu Phe
               1175                1180                1185
Pro Glu Ala Asn Phe Glu Leu Ser Ala Thr Asp Phe Leu Val Cys
               1190                1195                1200
Phe Ile Pro Tyr Phe Gln Thr Leu Leu Phe Val Phe Val Leu Arg
```

```
             1205                1210                1215
Cys Met Glu Leu Lys Cys Gly Lys Lys Arg Met Arg Lys Asp Pro
             1220                1225                1230
Val Phe Arg Ile Ser Pro Gln Ser Arg Asp Ala Lys Pro Asn Pro
             1235                1240                1245
Glu Glu Pro Ile Asp Glu Asp Glu Asp Ile Gln Thr Glu Arg Ile
             1250                1255                1260
Arg Thr Ala Thr Ala Leu Thr Thr Ser Ile Leu Asp Glu Lys Pro
             1265                1270                1275
Val Ile Ile Ala Ser Cys Leu His Lys Glu Tyr Ala Gly Gln Lys
             1280                1285                1290
Lys Ser Cys Phe Ser Lys Arg Lys Lys Lys Ile Ala Ala Arg Asn
             1295                1300                1305
Ile Ser Phe Cys Val Gln Glu Gly Glu Ile Leu Gly Leu Leu Gly
             1310                1315                1320
Pro Ser Gly Ala Gly Lys Ser Ser Ser Ile Arg Met Ile Ser Gly
             1325                1330                1335
Ile Thr Lys Pro Thr Ala Gly Glu Val Glu Leu Lys Gly Cys Ser
             1340                1345                1350
Ser Val Leu Gly His Leu Gly Tyr Cys Pro Gln Glu Asn Val Leu
             1355                1360                1365
Trp Pro Met Leu Thr Leu Arg Glu His Leu Glu Val Tyr Ala Ala
             1370                1375                1380
Val Lys Gly Leu Arg Lys Ala Asp Ala Arg Leu Ala Ile Ala Arg
             1385                1390                1395
Leu Val Ser Ala Phe Lys Leu His Glu Gln Leu Asn Val Pro Val
             1400                1405                1410
Gln Lys Leu Thr Ala Gly Ile Thr Arg Lys Leu Cys Phe Val Leu
             1415                1420                1425
Ser Leu Leu Gly Asn Ser Pro Val Leu Leu Leu Asp Glu Pro Ser
             1430                1435                1440
Thr Gly Ile Asp Pro Thr Gly Gln Gln Gln Met Trp Gln Ala Ile
             1445                1450                1455
Gln Ala Val Val Lys Asn Thr Glu Arg Gly Val Leu Leu Thr Thr
             1460                1465                1470
His Asn Leu Ala Glu Ala Glu Ala Leu Cys Asp Arg Val Ala Ile
             1475                1480                1485
Met Val Ser Gly Arg Leu Arg Cys Ile Gly Ser Ile Gln His Leu
             1490                1495                1500
Lys Asn Lys Leu Gly Lys Asp Tyr Ile Leu Glu Leu Lys Val Lys
             1505                1510                1515
Glu Thr Ser Gln Val Thr Leu Val His Thr Glu Ile Leu Lys Leu
             1520                1525                1530
Phe Pro Gln Ala Ala Gly Gln Glu Arg Tyr Ser Ser Leu Leu Thr
             1535                1540                1545
Tyr Lys Leu Pro Val Ala Asp Val Tyr Pro Leu Ser Gln Thr Phe
             1550                1555                1560
His Lys Leu Glu Ala Val Lys His Asn Phe Asn Leu Glu Glu Tyr
             1565                1570                1575
Ser Leu Ser Gln Cys Thr Leu Glu Lys Val Phe Leu Glu Leu Ser
             1580                1585                1590
Lys Glu Gln Glu Val Gly Asn Phe Asp Glu Glu Ile Asp Thr Thr
             1595                1600                1605
Met Arg Trp Lys Leu Leu Pro His Ser Asp Glu Pro
             1610                1615
```


<210> 17
<211> 1192
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6427133CD1

<400> 17

```
Met Phe Cys Ser Glu Lys Lys Leu Arg Glu Val Glu Arg Ile Val
  1               5                  10                  15
Lys Ala Asn Asp Arg Glu Tyr Asn Glu Lys Phe Gln Tyr Ala Asp
                 20                  25                  30
Asn Arg Ile His Thr Ser Lys Tyr Asn Ile Leu Thr Phe Leu Pro
                 35                  40                  45
Ile Asn Leu Phe Glu Gln Phe Gln Arg Val Ala Asn Ala Tyr Phe
                 50                  55                  60
Leu Cys Leu Leu Ile Leu Gln Leu Ile Pro Glu Ile Ser Ser Leu
                 65                  70                  75
Thr Trp Phe Thr Thr Ile Val Pro Leu Val Leu Val Ile Thr Met
                 80                  85                  90
Thr Ala Val Lys Asp Ala Thr Asp Asp Tyr Phe Arg His Lys Ser
                 95                 100                 105
Asp Asn Gln Val Asn Asn Arg Gln Ser Glu Val Leu Ile Asn Ser
                110                 115                 120
Lys Leu Gln Asn Glu Lys Trp Met Asn Val Lys Val Gly Asp Ile
                125                 130                 135
Ile Lys Leu Glu Asn Asn Gln Phe Val Ala Ala Asp Leu Leu Leu
                140                 145                 150
Leu Ser Ser Ser Glu Pro His Gly Leu Cys Tyr Val Glu Thr Ala
                155                 160                 165
Glu Leu Asp Gly Glu Thr Asn Leu Lys Val Arg His Ala Leu Ser
                170                 175                 180
Val Thr Ser Glu Leu Gly Ala Asp Ile Ser Arg Leu Ala Gly Phe
                185                 190                 195
Asp Gly Ile Val Val Cys Glu Val Pro Asn Asn Lys Leu Asp Lys
                200                 205                 210
Phe Met Gly Ile Leu Ser Trp Lys Asp Ser Lys His Ser Leu Asn
                215                 220                 225
Asn Glu Lys Ile Ile Pro Arg Gly Cys Ile Leu Arg Asn Thr Ser
                230                 235                 240
Trp Cys Phe Gly Met Val Ile Phe Ala Gly Pro Asp Thr Lys Leu
                245                 250                 255
Met Gln Asn Ser Gly Lys Thr Lys Phe Lys Arg Thr Ser Ile Asp
                260                 265                 270
Arg Leu Met Asn Thr Leu Val Leu Trp Ile Phe Gly Phe Leu Ile
                275                 280                 285
Cys Leu Gly Ile Ile Leu Ala Ile Gly Asn Ser Ile Trp Glu Ser
                290                 295                 300
Gln Thr Gly Asp Gln Phe Arg Thr Phe Leu Phe Trp Asn Glu Gly
                305                 310                 315
Glu Lys Ser Ser Val Phe Ser Gly Phe Leu Thr Phe Trp Ser Tyr
                320                 325                 330
Ile Ile Ile Leu Asn Thr Val Val Pro Ile Ser Leu Tyr Val Ser
                335                 340                 345
```

```
Val Glu Val Ile Arg Leu Gly His Ser Tyr Phe Ile Asn Trp Asp
                350                 355                 360
Arg Lys Met Tyr Tyr Ser Arg Lys Ala Ile Pro Ala Val Ala Arg
                365                 370                 375
Thr Thr Thr Leu Asn Glu Glu Leu Gly Gln Ile Glu Tyr Ile Phe
                380                 385                 390
Ser Asp Lys Thr Gly Thr Leu Thr Gln Asn Ile Met Thr Phe Lys
                395                 400                 405
Arg Cys Ser Ile Asn Gly Arg Ile Tyr Gly Glu Val His Asp Asp
                410                 415                 420
Leu Asp Gln Lys Thr Glu Ile Thr Gln Glu Lys Glu Pro Val Asp
                425                 430                 435
Phe Ser Val Lys Ser Gln Ala Asp Arg Glu Phe Gln Phe Phe Asp
                440                 445                 450
His Asn Leu Met Glu Ser Ile Lys Met Gly Asp Pro Lys Val His
                455                 460                 465
Glu Phe Leu Arg Leu Leu Ala Leu Cys His Thr Val Met Ser Glu
                470                 475                 480
Glu Asn Ser Ala Gly Glu Leu Ile Tyr Gln Val Gln Ser Pro Asp
                485                 490                 495
Glu Gly Ala Leu Val Thr Ala Ala Arg Asn Phe Gly Phe Ile Phe
                500                 505                 510
Lys Ser Arg Thr Pro Glu Thr Ile Thr Ile Glu Glu Leu Gly Thr
                515                 520                 525
Leu Val Thr Tyr Gln Leu Leu Ala Phe Leu Asp Phe Asn Asn Thr
                530                 535                 540
Arg Lys Arg Met Ser Val Ile Val Arg Asn Pro Glu Gly Gln Ile
                545                 550                 555
Lys Leu Tyr Ser Lys Gly Ala Asp Thr Ile Leu Phe Glu Lys Leu
                560                 565                 570
His Pro Ser Asn Glu Val Leu Leu Ser Leu Thr Ser Asp His Leu
                575                 580                 585
Ser Glu Phe Ala Gly Glu Gly Leu Arg Thr Leu Ala Ile Ala Tyr
                590                 595                 600
Arg Asp Leu Asp Asp Lys Tyr Phe Lys Glu Trp His Lys Met Leu
                605                 610                 615
Glu Asp Ala Asn Ala Ala Thr Glu Glu Arg Asp Glu Arg Ile Ala
                620                 625                 630
Gly Leu Tyr Glu Glu Ile Glu Arg Asp Leu Met Leu Leu Gly Ala
                635                 640                 645
Thr Ala Val Glu Asp Lys Leu Gln Glu Gly Val Ile Glu Thr Val
                650                 655                 660
Thr Ser Leu Ser Leu Ala Asn Ile Lys Ile Trp Val Leu Thr Gly
                665                 670                 675
Asp Lys Gln Glu Thr Ala Ile Asn Ile Gly Tyr Ala Cys Asn Met
                680                 685                 690
Leu Thr Asp Asp Met Asn Asp Val Phe Val Ile Ala Gly Asn Asn
                695                 700                 705
Ala Val Glu Val Arg Glu Glu Leu Arg Lys Ala Lys Gln Asn Leu
                710                 715                 720
Phe Gly Gln Asn Arg Asn Phe Ser Asn Gly His Val Val Cys Glu
                725                 730                 735
Lys Lys Gln Gln Leu Glu Leu Asp Ser Ile Val Glu Glu Thr Ile
                740                 745                 750
Thr Gly Asp Tyr Ala Leu Ile Ile Asn Gly His Ser Leu Ala His
                755                 760                 765
```



```
Ala Leu Glu Ser Asp Val Lys Asn Asp Leu Leu Glu Leu Ala Cys
                770                 775                 780
Met Cys Lys Thr Val Ile Cys Cys Arg Val Thr Pro Leu Gln Lys
                785                 790                 795
Ala Gln Val Val Glu Leu Val Lys Lys Tyr Arg Asn Ala Val Thr
                800                 805                 810
Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Ser Met Ile Lys Ser
                815                 820                 825
Ala His Ile Gly Val Gly Ile Ser Gly Gln Glu Gly Leu Gln Ala
                830                 835                 840
Val Leu Ala Ser Asp Tyr Ser Phe Ala Gln Phe Arg Tyr Leu Gln
                845                 850                 855
Arg Leu Leu Leu Val His Gly Arg Trp Ser Tyr Phe Arg Met Cys
                860                 865                 870
Lys Phe Leu Cys Tyr Phe Phe Tyr Lys Asn Phe Ala Phe Thr Leu
                875                 880                 885
Val His Phe Trp Phe Gly Phe Phe Cys Gly Phe Ser Ala Gln Thr
                890                 895                 900
Val Tyr Asp Gln Trp Phe Ile Thr Leu Phe Asn Ile Val Tyr Thr
                905                 910                 915
Ser Leu Pro Val Leu Ala Met Gly Ile Phe Asp Gln Asp Val Ser
                920                 925                 930
Asp Gln Asn Ser Val Asp Cys Pro Gln Leu Tyr Lys Pro Gly Gln
                935                 940                 945
Leu Asn Leu Leu Phe Asn Lys Arg Lys Phe Phe Ile Cys Val Leu
                950                 955                 960
His Gly Ile Tyr Thr Ser Leu Val Leu Phe Phe Ile Pro Tyr Gly
                965                 970                 975
Ala Phe Tyr Asn Val Ala Gly Glu Asp Gly Gln His Ile Ala Asp
                980                 985                 990
Tyr Gln Ser Phe Ala Val Thr Met Ala Thr Ser Leu Val Ile Val
                995                1000                1005
Val Ser Val Gln Ile Ala Leu Asp Thr Ser Tyr Trp Thr Phe Ile
               1010                1015                1020
Asn His Val Phe Ile Trp Gly Ser Ile Ala Ile Tyr Phe Ser Ile
               1025                1030                1035
Leu Phe Thr Met His Ser Asn Gly Ile Phe Gly Ile Phe Pro Asn
               1040                1045                1050
Gln Phe Pro Phe Val Gly Asn Ala Arg His Ser Leu Thr Gln Lys
               1055                1060                1065
Cys Ile Trp Leu Val Ile Leu Leu Thr Thr Val Ala Ser Val Met
               1070                1075                1080
Pro Val Val Ala Phe Arg Phe Leu Lys Val Asp Leu Tyr Pro Thr
               1085                1090                1095
Leu Ser Asp Gln Ile Arg Arg Trp Gln Lys Ala Gln Lys Lys Ala
               1100                1105                1110
Arg Pro Pro Ser Ser Arg Arg Pro Arg Thr Arg Arg Ser Ser Ser
               1115                1120                1125
Arg Arg Ser Gly Tyr Ala Phe Ala His Gln Glu Gly Tyr Gly Glu
               1130                1135                1140
Leu Ile Thr Ser Gly Lys Asn Met Arg Ala Lys Asn Pro Pro Pro
               1145                1150                1155
Thr Ser Gly Leu Glu Lys Thr His Tyr Asn Ser Thr Ser Trp Ile
               1160                1165                1170
Glu Asn Leu Cys Lys Lys Thr Thr Asp Thr Val Ser Ser Phe Ser
               1175                1180                1185
```

```
Gln Asp Lys Thr Val Lys Leu
            1190

<210> 18
<211> 625
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472932CD1

<400> 18
Met Ala His Ala Pro Glu Pro Asp Pro Ala Ala Ser Asp Leu Gly
  1               5                  10                  15
Asp Glu Arg Pro Lys Trp Asp Asn Lys Ala Gln Tyr Leu Leu Ser
             20                  25                  30
Cys Ile Gly Phe Ala Val Gly Leu Gly Asn Ile Trp Arg Phe Pro
             35                  40                  45
Tyr Leu Cys Gln Thr Tyr Gly Gly Gly Ala Phe Leu Ile Pro Tyr
             50                  55                  60
Val Ile Ala Leu Val Phe Glu Gly Ile Pro Ile Phe His Val Glu
             65                  70                  75
Leu Ala Ile Gly Gln Arg Leu Arg Lys Gly Ser Val Gly Val Trp
             80                  85                  90
Thr Ala Ile Ser Pro Tyr Leu Ser Gly Val Gly Leu Gly Cys Val
             95                 100                 105
Thr Leu Ser Phe Leu Ile Ser Leu Tyr Tyr Asn Thr Ile Val Ala
            110                 115                 120
Trp Val Leu Trp Tyr Leu Leu Asn Ser Phe Gln His Pro Leu Pro
            125                 130                 135
Trp Ser Ser Cys Pro Pro Asp Leu Asn Arg Thr Gly Phe Val Glu
            140                 145                 150
Glu Cys Gln Gly Ser Ser Ala Val Ser Tyr Phe Trp Tyr Arg Gln
            155                 160                 165
Thr Leu Asn Ile Thr Ala Asp Ile Asn Asp Ser Gly Ser Ile Gln
            170                 175                 180
Trp Trp Leu Leu Ile Cys Leu Ala Ala Ser Trp Ala Val Val Tyr
            185                 190                 195
Met Cys Val Ile Arg Gly Ile Glu Thr Thr Gly Lys Val Ile Tyr
            200                 205                 210
Phe Thr Ala Leu Phe Pro Tyr Leu Val Leu Thr Ile Phe Leu Ile
            215                 220                 225
Arg Gly Leu Thr Leu Pro Gly Ala Thr Lys Gly Leu Ile Tyr Leu
            230                 235                 240
Phe Thr Pro Asn Met His Ile Leu Gln Asn Pro Arg Val Trp Leu
            245                 250                 255
Asp Ala Ala Thr Gln Ile Phe Phe Ser Leu Ser Leu Ala Phe Gly
            260                 265                 270
Gly His Ile Ala Phe Ala Ser Tyr Asn Ser Pro Arg Asn Asp Cys
            275                 280                 285
Gln Lys Asp Ala Val Val Ile Ala Leu Val Asn Arg Met Thr Ser
            290                 295                 300
Leu Tyr Ala Ser Ile Ala Val Phe Ser Val Leu Gly Phe Lys Ala
            305                 310                 315
Thr Asn Asp Cys Pro Arg Arg Asn Ile Leu Ser Leu Ile Asn Asp
```



```
            320                 325                 330
Phe Asp Phe Pro Glu Gln Ser Ile Ser Arg Asp Asp Tyr Pro Ala
            335                 340                 345
Val Leu Met His Leu Asn Ala Thr Trp Pro Lys Arg Val Ala Gln
            350                 355                 360
Leu Pro Leu Lys Ala Cys Leu Leu Glu Asp Phe Leu Asp Lys Ser
            365                 370                 375
Ala Ser Gly Pro Gly Leu Ala Phe Val Val Phe Thr Glu Thr Asp
            380                 385                 390
Leu His Met Pro Gly Ala Pro Val Trp Ala Met Leu Phe Phe Gly
            395                 400                 405
Met Leu Phe Thr Leu Gly Leu Ser Thr Met Phe Gly Thr Val Glu
            410                 415                 420
Ala Val Ile Thr Pro Leu Leu Asp Val Gly Val Leu Pro Arg Trp
            425                 430                 435
Val Pro Lys Glu Ala Leu Thr Gly Leu Val Cys Leu Val Cys Phe
            440                 445                 450
Leu Ser Ala Thr Cys Phe Thr Leu Gln Ser Gly Asn Tyr Trp Leu
            455                 460                 465
Glu Ile Phe Asp Asn Phe Ala Ala Ser Leu Asn Leu Leu Met Leu
            470                 475                 480
Ala Phe Leu Glu Val Val Gly Val Val Tyr Val Tyr Gly Met Lys
            485                 490                 495
Arg Phe Cys Asp Asp Ile Ala Trp Met Thr Gly Arg Arg Pro Ser
            500                 505                 510
Pro Tyr Trp Arg Leu Thr Trp Arg Val Val Ser Pro Leu Leu Leu
            515                 520                 525
Thr Ile Phe Val Ala Tyr Ile Ile Leu Leu Phe Trp Lys Pro Leu
            530                 535                 540
Arg Tyr Lys Ala Trp Asn Pro Lys Tyr Glu Leu Phe Pro Ser Arg
            545                 550                 555
Gln Glu Lys Leu Tyr Pro Gly Trp Ala Arg Ala Ala Cys Val Leu
            560                 565                 570
Leu Ser Leu Leu Pro Val Leu Trp Val Pro Val Ala Ala Leu Ala
            575                 580                 585
Gln Leu Leu Thr Arg Arg Arg Arg Thr Trp Arg Asp Arg Asp Ala
            590                 595                 600
Arg Pro Asp Thr Asp Met Arg Pro Asp Thr Asp Thr Arg Pro Asp
            605                 610                 615
Thr Asp Met Arg Pro Asp Thr Asp Met Arg
            620                 625

<210> 19
<211> 1181
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8463147CD1

<400> 19
Met Thr Gln Ala Tyr Gln Lys Tyr Ile Leu Glu Lys Leu Pro Lys
  1               5                  10                  15
Ser Pro Gly Asp Lys Gly Arg Ala Trp Pro Gly Ser Thr Pro Ser
             20                  25                  30
```

```
Gly Asn Leu Leu Ser Pro Phe Met Ala Ala Ser Asn Ser Phe Pro
                35                  40                  45
Glu Leu Cys Ser Gln Val Ser Arg Arg Glu Tyr Trp Asp Leu His
            50                  55                  60
Gly Ile Pro Ser Asp His Phe Ser Val Arg Val Gln Val Glu Phe
        65                  70                  75
Tyr Met Asn Glu Asn Thr Phe Lys Glu Arg Leu Thr Leu Phe Phe
    80                  85                  90
Ile Thr Asn Gln Arg Ser Ser Leu Arg Ile Arg Leu Phe Asn Phe
                95                  100                 105
Ser Leu Lys Leu Leu Ser Cys Leu Leu Tyr Ile Ile Arg Val Leu
            110                 115                 120
Leu Glu Asn Pro Ser Gln Gly Asn Glu Trp Ser His Ile Phe Trp
        125                 130                 135
Val Asn Arg Ser Leu Pro Leu Trp Gly Leu Gln Val Ser Val Ala
    140                 145                 150
Leu Ile Ser Leu Phe Glu Thr Ile Leu Leu Gly Tyr Leu Ser Tyr
155                 160                 165
Lys Gly Asn Ile Trp Glu Gln Ile Leu Arg Ile Pro Phe Ile Leu
                170                 175                 180
Glu Ile Ile Asn Ala Val Pro Phe Ile Ile Ser Ile Phe Trp Pro
            185                 190                 195
Ser Leu Arg Asn Leu Phe Val Pro Val Phe Leu Asn Cys Trp Leu
        200                 205                 210
Ala Lys His Ala Leu Glu Asn Met Ile Asn Asp Leu His Arg Ala
    215                 220                 225
Ile Gln Arg Thr Gln Cys Cys Lys Cys Val Asn Gln Val Leu Ile
230                 235                 240
Val Ile Ser Thr Leu Leu Cys Leu Ile Phe Thr Cys Ile Cys Gly
                245                 250                 255
Ile Gln His Leu Glu Arg Ile Gly Lys Lys Leu Asn Leu Phe Asp
            260                 265                 270
Ser Leu Tyr Phe Cys Ile Val Thr Phe Ser Thr Val Gly Phe Gly
        275                 280                 285
Asp Val Thr Pro Glu Thr Trp Ser Ser Lys Leu Phe Val Val Ala
    290                 295                 300
Met Ile Cys Val Ala Leu Val Val Leu Pro Ile Gln Phe Glu Gln
305                 310                 315
Leu Ala Tyr Leu Trp Met Glu Arg Gln Lys Ser Gly Gly Asn Tyr
                320                 325                 330
Ser Arg His Arg Ala Gln Thr Glu Lys His Val Val Leu Cys Val
            335                 340                 345
Ser Ser Leu Lys Ile Asp Leu Leu Met Asp Phe Leu Asn Glu Phe
        350                 355                 360
Tyr Ala His Pro Arg Leu Gln Asp Tyr Tyr Val Val Ile Leu Cys
    365                 370                 375
Pro Thr Glu Met Asp Val Gln Val Arg Arg Val Leu Gln Ile Pro
380                 385                 390
Met Trp Ser Gln Arg Val Ile Tyr Leu Gln Gly Ser Ala Leu Lys
                395                 400                 405
Asp Gln Asp Leu Leu Arg Ala Lys Met Asp Asp Ala Glu Ala Cys
            410                 415                 420
Phe Ile Leu Ser Ser Arg Cys Glu Val Asp Arg Thr Ser Ser Asp
        425                 430                 435
His Gln Thr Ile Leu Arg Ala Trp Ala Val Lys Asp Phe Ala Pro
    440                 445                 450
```



```
Asn Cys Pro Leu Tyr Val Gln Ile Leu Lys Pro Glu Asn Lys Phe
                455                 460                 465
His Ile Lys Phe Ala Asp His Val Val Cys Glu Glu Glu Phe Lys
            470                 475                 480
Tyr Ala Met Leu Ala Leu Asn Cys Ile Cys Pro Ala Thr Ser Thr
        485                 490                 495
Leu Ile Thr Leu Leu Val His Thr Ser Arg Gly Gln Cys Val Cys
    500                 505                 510
Leu Cys Cys Arg Glu Gly Gln Gln Ser Pro Glu Gln Trp Gln Lys
515                 520                 525
Met Tyr Gly Arg Cys Ser Gly Asn Glu Val Tyr His Ile Val Leu
                530                 535                 540
Glu Glu Ser Thr Phe Phe Ala Glu Tyr Glu Gly Lys Ser Phe Thr
            545                 550                 555
Tyr Ala Ser Phe His Ala His Lys Lys Phe Gly Val Cys Leu Ile
        560                 565                 570
Gly Val Arg Arg Glu Asp Asn Lys Asn Ile Leu Leu Asn Pro Gly
    575                 580                 585
Pro Arg Tyr Ile Met Asn Ser Thr Asp Ile Cys Phe Tyr Ile Asn
590                 595                 600
Ile Thr Lys Glu Glu Asn Ser Ala Phe Lys Asn Gln Asp Gln Gln
                605                 610                 615
Arg Lys Ser Asn Val Ser Arg Ser Phe Tyr His Gly Pro Ser Arg
            620                 625                 630
Leu Pro Val His Ser Ile Ile Ala Ser Met Gly Thr Val Ala Ile
        635                 640                 645
Asp Leu Gln Asp Thr Ser Cys Arg Ser Ala Ser Gly Pro Thr Leu
    650                 655                 660
Ser Leu Pro Thr Glu Gly Ser Lys Glu Ile Arg Arg Pro Ser Ile
665                 670                 675
Ala Pro Val Leu Glu Val Ala Asp Thr Ser Ser Ile Gln Thr Cys
                680                 685                 690
Asp Leu Leu Ser Asp Gln Ser Glu Asp Glu Thr Thr Pro Asp Glu
            695                 700                 705
Glu Met Ser Ser Asn Leu Glu Tyr Ala Lys Gly Tyr Pro Pro Tyr
        710                 715                 720
Ser Pro Tyr Ile Gly Ser Ser Pro Thr Phe Cys His Leu Leu His
    725                 730                 735
Glu Lys Val Pro Phe Cys Cys Leu Arg Leu Asp Lys Ser Cys Gln
740                 745                 750
His Asn Tyr Tyr Glu Asp Ala Lys Ala Tyr Gly Phe Lys Asn Lys
                755                 760                 765
Leu Ile Ile Val Ala Ala Glu Thr Ala Gly Asn Gly Leu Tyr Asn
            770                 775                 780
Phe Ile Val Pro Leu Arg Ala Tyr Tyr Arg Pro Lys Lys Glu Leu
        785                 790                 795
Asn Pro Ile Val Leu Leu Leu Asp Asn Pro Pro Asp Met His Phe
    800                 805                 810
Leu Asp Ala Ile Cys Trp Phe Pro Met Val Tyr Tyr Met Val Gly
815                 820                 825
Ser Ile Asp Asn Leu Asp Asp Leu Leu Arg Cys Gly Val Thr Phe
                830                 835                 840
Ala Ala Asn Met Val Val Val Asp Lys Glu Ser Thr Met Ser Ala
            845                 850                 855
Glu Glu Asp Tyr Met Ala Asp Ala Lys Thr Ile Val Asn Val Gln
        860                 865                 870
```

```
Thr Leu Phe Arg Leu Phe Ser Ser Leu Ser Ile Ile Thr Glu Leu
                875                 880                 885
Thr His Pro Ala Asn Met Arg Phe Met Gln Phe Arg Ala Lys Asp
                890                 895                 900
Cys Tyr Ser Leu Ala Leu Ser Lys Leu Glu Lys Lys Glu Arg Glu
                905                 910                 915
Arg Gly Ser Asn Leu Ala Phe Met Phe Arg Leu Pro Phe Ala Ala
                920                 925                 930
Gly Arg Val Phe Ser Ile Ser Met Leu Asp Thr Leu Leu Tyr Gln
                935                 940                 945
Ser Phe Val Lys Asp Tyr Met Ile Ser Ile Thr Arg Leu Leu Leu
                950                 955                 960
Gly Leu Asp Thr Thr Pro Gly Ser Gly Phe Leu Cys Ser Met Lys
                965                 970                 975
Ile Thr Ala Asp Asp Leu Trp Ile Arg Thr Tyr Ala Arg Leu Tyr
                980                 985                 990
Gln Lys Leu Cys Ser Ser Thr Gly Asp Val Pro Ile Gly Ile Tyr
                995                1000                1005
Arg Thr Glu Ser Gln Lys Leu Thr Thr Ser Glu Ser Gln Ile Ser
            1010                1015                1020
Ile Ser Val Glu Glu Trp Glu Asp Thr Lys Asp Ser Lys Glu Gln
            1025                1030                1035
Gly His His Arg Ser Asn His Arg Asn Ser Thr Ser Ser Asp Gln
            1040                1045                1050
Ser Asp His Pro Leu Leu Arg Arg Lys Ser Met Gln Trp Ala Arg
            1055                1060                1065
Arg Leu Ser Arg Lys Gly Pro Lys His Ser Gly Lys Thr Ala Glu
            1070                1075                1080
Lys Ile Thr Gln Gln Arg Leu Asn Leu Tyr Arg Arg Ser Glu Arg
            1085                1090                1095
Gln Glu Leu Ala Glu Leu Val Lys Asn Arg Met Lys His Leu Gly
            1100                1105                1110
Leu Ser Thr Val Gly Tyr Asp Glu Met Asn Asp His Gln Ser Thr
            1115                1120                1125
Leu Ser Tyr Ile Leu Ile Asn Pro Ser Pro Asp Thr Arg Ile Glu
            1130                1135                1140
Leu Asn Asp Val Val Tyr Leu Ile Arg Pro Asp Pro Leu Ala Tyr
            1145                1150                1155
Leu Pro Asn Ser Glu Pro Ser Arg Arg Asn Ser Ile Cys Asn Val
            1160                1165                1170
Thr Gly Gln Asp Ser Arg Glu Glu Thr Gln Leu
            1175                1180
```

<210> 20
<211> 233
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506408CD1

<400> 20

```
Met Leu Glu Gly Ala Glu Leu Tyr Phe Asn Val Asp His Gly Tyr
  1               5                  10                  15
Leu Glu Gly Leu Val Arg Gly Cys Lys Ala Ser Leu Leu Thr Gln
```



```
                20                  25                  30
Gln Asp Tyr Ile Asn Leu Val Gln Cys Glu Thr Leu Glu Ala Pro
                35                  40                  45
Phe Phe Gln Asp Cys Met Ser Glu Asn Ala Leu Asp Glu Leu Asn
                50                  55                  60
Ile Glu Leu Leu Arg Asn Lys Leu Tyr Lys Ser Tyr Leu Glu Ala
                65                  70                  75
Phe Tyr Lys Phe Cys Lys Asn His Gly Asp Val Thr Ala Glu Val
                80                  85                  90
Met Cys Pro Ile Leu Glu Phe Glu Ala Asp Arg Arg Ala Phe Ile
                95                 100                 105
Ile Thr Leu Asn Ser Phe Gly Thr Glu Leu Ser Lys Glu Asp Arg
               110                 115                 120
Glu Thr Leu Tyr Pro Thr Phe Gly Lys Leu Tyr Pro Glu Gly Leu
               125                 130                 135
Arg Leu Leu Ala Gln Ala Glu Asp Phe Asp Gln Met Lys Asn Val
               140                 145                 150
Ala Asp His Tyr Gly Val Tyr Lys Pro Leu Phe Glu Ala Val Gly
               155                 160                 165
Gly Ser Gly Gly Lys Thr Leu Glu Asp Val Phe Tyr Glu Arg Glu
               170                 175                 180
Val Gln Met Asn Val Leu Ala Phe Asn Arg Gln Phe His Tyr Gly
               185                 190                 195
Val Phe Tyr Ala Tyr Val Lys Leu Lys Glu Gln Glu Ile Arg Asn
               200                 205                 210
Ile Val Trp Ile Ala Glu Cys Ile Ser Gln Arg His Arg Thr Lys
               215                 220                 225
Ile Asn Ser Tyr Ile Pro Ile Leu
               230
```

<210> 21
<211> 2232
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6911460CB1

<400> 21

```
attagctttg cccgaagttt ttccccacac tcttctttag catgctatta tggggaaagt 60
gaccactcct gggagcgggg gtggtcgggg cggtttggtg gcggggaagc ggctgtaact 120
tctacgtgac catggtacct gttgaaaaca ccgagggccc cagtctgctg aaccagaagg 180
ggacagccgt ggagacggag ggcagcggca gccggcatcc tccctgggcg agaggctgcg 240
gcatgtttac cttcctgtca tctgtcactg ctgctgtcag tggcctcctg gtgggttatg 300
aacttgggat catctctggg gctcttcttc agatcaaaac cttattagcc ctgagctgcc 360
atgagcagga aatggttgtg agctccctcg tcattggagc cctccttgcc tcactcaccg 420
gaggggtcct gatagacaga tatggaagaa ggacagcaat catcttgtca tcctgcctgc 480
ttggactcgg aagcttagtc ttgatcctca gtttatccta cacggttctt atagtgggac 540
gcattgccat aggggtctcc atctccctct cttccattgc cacttgtgtt tacatcgcag 600
agattgctcc tcaacacaga agaggccttc ttgtgtcact gaatgagctg atgattgtca 660
tcggcattct ttctgcctat atttcaaatt acgcatttgc caatgttttc catggctgga 720
agtacatgtt tggtcttgtg attcccttgg gagttttgca agcaattgca atgtattttc 780
ttcctccaag ccctcggttt ctggtgatga aaggacaaga gggagctgct agcaaggttc 840
ttggaaggtt aagagcactc tcagatacaa ctgaggaact cactgtgatc aaatcctccc 900
tgaaagatga atatcagtac agtttttggg atctgtttcg ttcaaaagac aacatgcgga 960
```

cccgaataat gataggacta acactagtat tttttgtaca aatcactggc caaccaaaca 1020
tattgttcta tgcatcaact gttttgaagt cagttggatt tcaaagcaat gaggcagcta 1080
gcctcgcctc cactggggtt ggagtcgtca aggtcattag caccatccct gccactcttc 1140
ttgtagacca tgtcggcagc aaaacattcc tctgcattgg ctcctctgtg atggcagctt 1200
cgttggtgac catgggcatc gtaaatctca acatccacat gaacttcacc catatctgca 1260
gaagccacaa ttctatcaac cagtccttgg atgagtctgt gatttatgga ccaggaaacc 1320
tgtcaaccaa caacaatact ctcagagacc acttcaaagg gatttcttcc catagcagaa 1380
gctcactcat gcccctgaga aatgatgtgg ataagagagg ggagacgacc tcagcatcct 1440
tgctaaatgc tggattaagc cacactgaat accagatagt cacagaccct ggggacgtcc 1500
cagctttttt gaaatggctg tccttagcca gcttgcttgt ttatgttgct gctttttcaa 1560
ttggtctagg accaatgccc tggctggtgc tcagcgagat ctttcctggt gggatcagag 1620
gacgagccat ggctttaact tctagcatga actggggcat caatctcctc atctcgctga 1680
cattttgac tgtaactgat cttattggcc tgccatgggt gtgctttata tatacaatca 1740
tgagtctagc atccctgctt tttgttgtta tgtttatacc tgagacaaag ggatgctctt 1800
tggaacaaat atcaatggag ctagcaaaag tgaactatgt gaaaaacaac atttgtttta 1860
tgagtcatca ccaagaagaa ttagtgccaa aacagcctca aaaaagaaaa ccccaggagc 1920
agctctttgga gtgtaacaag ctgtgtggta ggggccaatc caggcagctt tctccagaga 1980
cctaatggcc tcaacacctt ctgaacgtgg atagtgccag aacacttagg agggtgtctt 2040
tggaccaatg catagttgcg actcctgtgc tctcttttca gtgtcatgga actggttttg 2100
aagagacact ctgaaatgat aaagacagcc tttaatcccc ctcctcccca gaaggaacct 2160
caaaaggtag atgaggtaca aggtcctaag tgatctcttt ttctgagcag gatatcaggt 2220
taaaaaaaaa aa 2232

<210> 22
<211> 4135
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55138203CB1

<400> 22
acaaccccac aggccagctt tttcacatag ttgttaccag cacttggcca acagttgttt 60
ttcatcagtg ggtggagcag cttttcttgc ccccaaaaaa cagtcaacca ctcatttttc 120
attgggtata tgtattcggc aaacattggg tacctgctgt ttgttggcac tggtgttgag 180
aagatgaata acacaccctc tatggcccta gggagttccc attctggtag ggggaacctg 240
actcaggcag caacaaaacc ttctggttat gagaagacag atgatgtttc agagaagacc 300
tcactggctg accaggagga agtaaggact attttcatca accagcccca gctgacaaaa 360
ttctgcaata accatgtcag cactgcaaaa tacaacataa tcacattcct tccaagattt 420
ctctactctc agttcagaag agctgctaat tcatttttc tctttattgc actgctgcag 480
caaatacctg atgtgtcacc aacaggtcgt tatacaacac tggttcctct cttatttatt 540
ttagctgtgg cagctatcaa agagataata gaagatatta aacgacataa agctgataat 600
gcagtgaaca agaaacaaac gcaagttttg agaaatggtg cttgggaaat tgtccactgg 660
gaaaaggtaa atgttggaga tatagttata ataaaaggca aagagtatat acctgctgac 720
actgtacttc tctcatcaag tgagccccaa gccatgtgct acattgaaac atccaactta 780
gatggtgaaa caaacttgaa aattagacag ggcttaccag caacatcaga tatcaaagac 840
gttgacagtt tgatgaggat ttctggcaga attgagtgtg aaagtccaaa cagacatctc 900
tacgattttg ttggaaacat aaggcttgat ggacatggca ccgttccact gggagcagat 960
cagattcttc ttcgaggagc tcagttgaga aatacacagt gggttcatgg aatagttgtc 1020
tacactggac atgacaccaa gctgatgcag aattcaacaa gtccaccact taagctctca 1080
aatgtggaac ggattacaaa tgtacaaatt ttgattttat tttgtatctt aattgccatg 1140
tctcttgtct gttctgtggg ctcagccatt tggaatcgaa ggcattctgg aaaagactgg 1200
tatctcaatc taaactatgg tggcgctagt aattttggac tgaatttctt gaccttcatc 1260
atcctttca acaatctcat tcctatcagc ttattggtta cattagaagt tgtgaaattt 1320
acccaggcat acttcataaa ttgggatctt gacatgcact atgaacccac agacactgct 1380

gctatggctc gaacatctaa tctgaatgag gaacttggcc aggttaaata catattttct 1440
gacaaaactg gtactctgac atgcaatgta atgcagttta agaagtgcac catagcggga 1500
gttgcttatg gccatgtccc tgaaccctgag gattatggct gctctcctga tgaatggcag 1560
aactcacagt ttggagatga aaaaacattt agtgattcat cattgctgga aaatctccaa 1620
aataatcatc ccaccgcacc tataatatgt gaatttctta caatgatggc agtctgtcac 1680
acagcagtgc cagagcgaga aggtgacaag attatttatc aagcagcatc tccagatgag 1740
ggagcattgg tcagagcagc caagcaattg aattttgttt tcactggaag aacacccgac 1800
tcggtgatta tagattcact ggggcaggaa gaaagatatg aattgctcaa tgtcttggag 1860
tttaccagtg ctaggaaaag aatgtcagtg attgttcgca ctccatctgg aaagttacga 1920
ctctactgca aaggagctga cactgtaatt tatgatcgac tggcagagac gtcaaaatac 1980
aaagaaatta ccctaaaaca tttagagcag tttgctacag aagggttaag aactttatgt 2040
tttgctgtgg ctgagatttc agagagcgac tttcaggagt ggcgagcagt ctatcagcga 2100
gcatctacat ctgtgcagaa caggctactc aaactcgaag agagttatga gttgattgaa 2160
aagaatcttc agctacttgg agcaacagcc attgaggata aattacaaga tcaagtgcct 2220
gaaaccatag aaacgctaat gaaagcagac atcaaaatct ggatccttac aggggacaag 2280
caagaaactg ccattaacat cggacactcc tgcaaactgt tgaagaagaa catgggaatg 2340
attgttataa atgaaggctc tcttgatgga acaagggaaa ctctcagtcg tcactgtact 2400
acccttggtg atgctctccg gaaagagaat gattttgctc ttataattga tgggaaaacc 2460
ctcaaatatg ccttaacctt tggagtacga cagtatttcc tggacttagc tttgtcatgc 2520
aaagctgtca tttgctgtcg ggtttctcct cttcaaaaat ctgaagttgt tgagatggtt 2580
aagaaacaag tcaaagtcgt aacgcttgca atcggtgatg gagcaaatga tgtcagcatg 2640
atacagacag cgcacgttgg tgttggtatc agtggcaatg aaggcctgca ggcagctaat 2700
tcctctgact actccatagc tcagttcaaa tatttgaaga atttactgat gattcatggt 2760
gcctggaact ataacagagt ctccaagtgc atcttatact gcttctacaa gaatatagtg 2820
ctctatatta tcgagatctg gtttgccttt gttaatggct tttctggaca gatcctcttt 2880
gaaagatggt gtataggtct ctataacgtg atgtttacag caatgcctcc tttaactctt 2940
ggaatatttg agagatcatg cagaaaagag aacatgttga agtaccctga attatacaaa 3000
acatctcaga atgccctgga cttcaacacc aaggttttct gggttcattg tttaaatggc 3060
ctcttccact cagttattct gttttggttt ccactaaaag cccttcagta tggtactgca 3120
tttggaaatg ggaaaacctc ggattatctg ctactgggaa actttgtgta cacttttgtg 3180
gtgataactg tgtgtttgaa agctggattg gagacatcat attggacatg gttcagccac 3240
atagcgatat gggggagcat cgcactctgg gtggtgtttt tgggaatcta ctcatctctg 3300
tggcctgcca ttccgatggc ccctgatatg tcaggagagg cagccatgtt gttcagttct 3360
ggagtctttt ggatgggctt gttattcatc cctgtggcat ctctgctcct tgatgtggtg 3420
tacaaggtta tcaagaggac tgcttttaaa acattggtcg atgaagttca ggagctggag 3480
gcaaaatctc aagacccagg agcagttgta cttggaaaaa gcctgaccga gagggcgcaa 3540
ctgctcaaga acgtctttaa gaagaaccac gtgaacttgt accgctctga atccttgcaa 3600
caaaatctgc tccatgggta tgcgttctct caagatgaaa atggaatcgt ttcacagtct 3660
gaagtgataa gagcatatga taccacgaaa cagaggcccg acgaatggtg atggggagag 3720
cctgaaaggc aggctctgtt acctctctaa ggagagctac caggttgtca ccgcagtctg 3780
ctaaccaatt ccagtctggt ccatgaagag gaaaggtaga tctgagctca tctcgctgat 3840
ggacattcag attcatgtat attatagaca taagcactgt gcaactgtac tgtaacacca 3900
tctcttttgg attttttaa ggtatttgct aagtctttgt aaacggaaat tgaaaatgac 3960
ctggtatctt gccagagggc tttcttaaac ggagaataag tcagtattct tatgccatta 4020
ctgtggggct gtaactgact gtcagtttat tggctgtacc acaaggtaac caaccattaa 4080
aaaactctaa atgatattta gttaaaggga ctctgtggta tccagactta gattt 4135

<210> 23
<211> 2970
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478871CB1

<400> 23
atgcaaccag ccagagggcc cctggcttca gaacctagga ctgtactggt tctgagattc 60
tgtgcaagcc tcatggaaat gaagctgcca ggccaggaag ggtttgaagc ctccagtgct 120
cctagaaata ttccttcagg ggagctggac agcaaccctg accctggcac cggccccagc 180
cctgatggcc cctcagacac agagagcaag gaactgggag tacccaaaga ccctctgctc 240
ttcattcagc tgaatgagct gctgggctgg ccccaggcgc tggagtggag agagacaggc 300
acgtgggtac tgtttgagga gaagttggag gtggctgcag gccggtggag tgccccccac 360
gtgcccaccc tggcactgcc cagcctccag aagctccgca gcctgctggc cgagggcctt 420
gtactgctgg actgcccagc tcagagcctc ctggagctcg tggagcaggt gaccagggtg 480
gagtcgctga gcccagagct gagagggcag ttgcaggcct tgctgctgca gagaccccag 540
cattacaacc agaccacagg caccaggccc tgctggggtg agagcccctc cctgggccca 600
ggaccaagac cctgtacaac cagaccacag gcaccaggcc ctgctgggca gtgtcagaac 660
cccctgagac agaagctacc tccaggagct gaggcaggga ctgtgctggc aggggagctg 720
ggcttcctgg cacagccact gggagccttt gttcgactgc ggaaccctgt ggtactgggg 780
tcccttactg aggtgtccct cccaagcagg tttttctgcc ttctcctggg cccctgtatg 840
ctgggaaagg gctaccatga gatgggacgg gcagcagctg tcctcctcag tgacccgcaa 900
ttccagtggt cagttcgtcg ggccagcaac cttcatgacc ttctggcagc cctggatgca 960
ttcctagagg aggtgacagt gcttccccca ggtcggtggg acccaacagc ccggattccc 1020
ccgcccaaat gtctgccatc tcagcacaaa aggcttccct cgcaacagcg ggagatcaga 1080
ggtcccgccg tcccgcgcct gacctcggct gaggacaggc accgccatgg gccacacgca 1140
cacagcccgg agttgcagcg gaccggcagg ctgtttgggg gccttatcca ggacgtgcgc 1200
aggaaggtcc cgtggtaccc cagcgatttc ttggacgccc tgcatctcca gtgcttctcg 1260
gccgtactct acatttacct ggccactgtc actaatgcca tcacttttgg gggtctgctg 1320
ggagatgcca ctgatggtgc ccagggagtg ctggaaagtt tcctgggcac agcagtggct 1380
ggagctgcct tctgcctgat ggcaggccag cccctcacca ttctgagcag cacggggcca 1440
gtgctggtct ttgagcgcct gctcttctct ttcagcagag attacagcct ggactacctg 1500
cccttccgcc tatgggtggg catctgggtg gctacctttt gcctggtgct ggtggccaca 1560
gaggccagtg tgctggtgcg ctacttcacc cgcttcactg aggaaggttt ctgtgccctc 1620
atcagcctca tcttcatcta cgatgctgtg ggcaaaatgc tgaacttgac ccatacctat 1680
cctatccaga agcctgggtc ctctgcctac gggtgcctct gccaataccc aggcccagga 1740
ggaaatgagt ctcaatggat aaggacaagg ccaaaagaca gagacgacat tgtaagcatg 1800
gacttaggcc tgatcaatgc atccttgctg ccgccacctg agtgcacccg gcagggaggc 1860
caccctcgtg gccctggctg tcatacagtc ccagacattg ccttcttctc ccttctcctc 1920
ttccttactt ctttcttctt tgctatggcc ctcaagtgtg taaagaccag ccgcttcttc 1980
ccctctgtgg tgcgcaaagg gctcagcgac ttctcctcag tcctggccat cctgctcggc 2040
tgtggccttg atgctttcct gggcctagcc acaccaaagc tcatggtacc cagagagttc 2100
aagcccacac tccctgggcg tggctggctg gtgtcacctt ttggagccaa cccctggtgg 2160
tggagtgtgg cagctgccct gcctgccctg ctgctgtcta tcctcatctt catggaccaa 2220
cagatcacag cagtcatcct caaccgcatg gaatacagac tgcagaaggg agctggcttc 2280
cacctggacc tcttctgtgt ggctgtgctg atgctactca catcagcgct tggactgcct 2340
tggtatgtct cagccactgt catctccctg gctcacatgg acagtcttcg gagagagagc 2400
agagcctgtg cccccgggga gcgccccaac ttcctgggta tcagggaaca gaggctgaca 2460
ggcctggtgg tgttcatcct tacaggagcc tccatcttcc tggcacctgt gctcaagttc 2520
attccaatgc ctgtgctcta tggcatcttc ctgtatatgg gggtggcagc gctcagcagc 2580
attcagttca ctaatagggt gaagctgttg ttgatgccag caaaacacca gccagacctg 2640
ctactcttgc ggcatgtgcc tctgaccagg gtccacctct tcacagccat ccagcttgcc 2700
tgtctggggc tgctttggat aatcaagtct acccctgcag ccatcatctt ccccctcatg 2760
ttgctgggcc ttgtgggggt ccgaaaggcc ctggagaggg tcttctcacc acaggaactc 2820
ctctggctgg atgagctgat gccagaggag gagagaagca tccctgagaa ggggctggag 2880
ccagaacact cattcagtgg aagtgacagt gaagattcag agctgatgta tcagccaaag 2940
gctccagaaa tcaacatttc tgtgaattag 2970

<210> 24
<211> 1835
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483601CB1

<400> 24
atggatcatg ctgaagaaaa tgaaatcctt gcagcaaccc agaggtacta tgtggaaagg 60
cctatcttta gtcatccggt cctccaggaa agactacaca caaaggacaa ggttcctgat 120
tccattgcgg ataagctgaa acaggcattc acatgtactc ctaaaaaaat aagaaatatc 180
atttatatgt tcctacccat aactaaatgg ctgccagcat acaaattcaa ggaatatgtg 240
ttgggtgact tggtctcagg cataagcaca ggggtgcttc agcttcctca aggcttagcc 300
tttgcaatgc tggcagctgt gcctccaata tttggcctgt acccttcatt ttaccctgtt 360
atcatgtatt gttttcttgg aacctccaga cacatatcca taggtccttt tgctgttatt 420
agcctgatga ttggtggtgt agctgttcga ttagtaccag atgatatagt cattccagga 480
ggagtaaatg caaccaatgg cacagaggcc agagatgcct tgagagtgaa agtcgccatg 540
tctgtgacct tactttcagg aatcattcag ttttgcctag gtgtctgtag gtttggattt 600
gtggccatat atctcacaga gcctctggtc cgtgggttta ccaccgcagc agctgtgcat 660
gtcttcacct ccatgttaaa atatctgttt ggagttaaaa caaagcggta cagtggaatc 720
ttttccgtgg tgtatagtac agttgctgtg ttgcagaatg ttaaaaacct caacgtgtgt 780
tccctaggcg tcgggctgat ggtttttggt ttgctgttgg gtggcaagga gtttaatgag 840
agatttaaag agaaattgcc ggcgcctatt cctttagagt tctttgcggt cgtaatggga 900
actggcattt cagctgggtt taacttgaaa gaatcataca atgtggatgt cgttggaaca 960
cttcctctag ggctgctacc tccagccaat ccggacacca gcctcttcca ccttgtgtac 1020
gtagatgcca ttgccatagc catcgttgga ttttcagtga ccatctccat ggccaagacc 1080
ttagcaaata aacatggcta ccaggttgac ggcaatcagg agctcattgc cctgggactg 1140
tgcaattcca ttggctcact cttccagacc ttttcaattt catgctcctt gtctcgaagc 1200
cttgttcagg agggaaccgg tgggaagaca cagcttgcag gttgtttggc ctcattaatg 1260
attctgctgg tcatattagc aactggattc ctctttgaat cattgcccca ggctgtgctg 1320
tcggccattg tgattgtcaa cctgaaggga atgtttatgc agttctcaga tctccccttt 1380
ttctggagaa ccagcaaaat agagctgacc atctggctta ccacttttgt gtcctccttg 1440
ttcctgggat tggactatgg tttgatcact gctgtgatca ttgctctgct gactgtgatt 1500
tacagaacac agaggtgaaa gaaattcctg gaataaaaat atttcaaata aatgccccaa 1560
tttactatgc aaatagggac tgtatagcca agcttaaaag aaagactggg gtgaacccag 1620
cagtcatcat ggggacaggg gaaaggcgtg gggaatacgc taagggagtc ggaatggaaa 1680
tgggcacggc atgtggtaag cgatgcggag tatgggggtt aacaagcgaa aaggggtgga 1740
gaaaattccc aatgtaaaag attttggaag gagaatgacc cggaagacac aagttgggtt 1800
tacaataggt tggggagacg gcggaaagag ggtta 1835

<210> 25
<211> 2220
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487851CB1

<400> 25
caaggcagca tgagccgatc acccctcaat cccagccaac tccgatcagt gggctcccag 60
gatgccctgg ccccttgcc tccacctgct ccccagaatc cctccaccca ctcttgggac 120
cctttgtgtg gatctctgcc ttggggcctc agctgtcttc tggctctgca gcatgtcttg 180
gtcatggctt ctctgctctg tgtctcccac ctgctcctgc tttgcagtct ctccccagga 240
ggactctctt actccccttc tcagctcctg gcctccagct tcttttcacg tggtatgtct 300
accatcctgc aaacttggat gggcagcagg ctgcctcttg tccaggctcc atccttagag 360
ttccttatcc ctgctctggt gctgaccagc cagaagctac cccgggccat ccagacacct 420
ggaaactgtg agcacagagc aagggcaagg gcctccctca tgctgcacct ttgtagggga 480
cctagctgcc atggcctggg gcactggaac acttctctcc aggaggtgtc cggggcagtg 540

gtagtatctg ggctgctgca gggcatgatg gggctgctgg ggagtcccgg ccacgtgttc 600
ccccactgtg ggcccctggt gctggctccc agcctggttg tggcagggct ctctgcccac 660
agggaggtag cccagttctg cttcacacac tgggggttgg ccttgctggt tatcctgctc 720
atggtggtct gttctcagca cctgggctcc tgccagtttc atgtgtgccc ctggaggcga 780
gcttcaacgt catcaactca cactcctctc cctgtcttcc ggctcctttc ggtgctgatc 840
ccagtggcct gtgtgtggat tgtttctgcc tttgtgggat tcagtgttat cccccaggaa 900
ctgtctgccc ccaccaaggc accatggatt tggctgcctc acccaggtga gtggaattgg 960
cctttgctga cgcccagagc tctggctgca ggcatctcca tggccttggc agcctccacc 1020
agttccctgg gctgctatgc cctgtgtggc cggctgctgc atttgcctcc cccacctcca 1080
catgcctgca gtcgagggct gagcctggag gggctgggca gtgtgctggc cgggctgctg 1140
ggaagcccca tgggcactgc atccagcttc cccaacgtgg gcaaagtggg tcttatccag 1200
gctggatctc agcaagtggc tcacttagtg gggctactct gcgtggggct tggactctcc 1260
cccaggttgg ctcagctcct caccaccatc ccactgcctg ttgttggtgg ggtgctgggg 1320
gtgacccagg ctgtggtttt gtctgctgga ttctccagct tctacctggc tgacatagac 1380
tctgggcgaa atatcttcat tgtgggcttc tccatcttca tggccttgct gctgccaaga 1440
tggtttcggg aagccccagt cctgttcagc acaggctgga gccccttgga tgtattactg 1500
cactcactgc tgacacagcc catcttcctg gctggactct caggcttcct actagagaac 1560
acgattcctg gcacacagct tgagcgaggc ctaggtcaag ggctaccatc tcctttcact 1620
gcccaagagg ctcgaatgcc tcagaagccc agggagaagg ctgctcaagt gtacagactt 1680
cctttcccca tccaaaacct ctgtccctgc atcccccagc ctctccactg cctctgccca 1740
ctgcctgaag accctgggga tgaggaagga ggctcctctg agccagaaga gatggcagac 1800
ttgctgcctg gctcagggga gccatgccct gaatctagca gagaagggtt taggtcccag 1860
aaatgaccag aacgcctact tctgccttgg ttaatttagc cctaactctc atctgctgga 1920
gagtcagctc ccaaactgtt ctttcttgta ggcagaggat atgtgtgtgt gtattacatg 1980
ggactgtcta gaggttccat ttcccaatag ggtgggttgc ctttccttgt cttaattagg 2040
cctaactgtt ccagagcaga ggccatgatt tagtggacca tgaatgattg agattttgcc 2100
tgtgtactat caatgccact tgaacccaag cattcacttt aatacttact gagcatctcc 2160
catgtgcaag gtcctggaac tacagggata agacagggtc catgccgtct caaggcattt 2220

<210> 26
<211> 1517
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472881CB1

<400> 26
taagaacaga agtggaaagc cttacttacc acagtttatt atatgtttca tgcccgtgat 60
aattactttt ataatgccac ttgtgaaaaa attgatcaga ttaggatgaa tcaccttgct 120
ggccaacagt tattggaatg attctccatg tgtgacttcg ttgcactatt acaaaatgtg 180
gcaggataga cctgcccagc cattgttgcc gatgttcatt tgtaatgctg ccttaaggag 240
atgaggagat gagagccaat tgttccagca gctcagcctg ccctgccaac agttcagagg 300
aggagctgcc agtgggactg gaggcgcatg gaaacctgga gctcgttttc acagtggtgc 360
ccactgtgat gatggggctg ctcatgttct ctttgggatg ttccgtggag atccggaagc 420
tgtggtcgca catcaggaga ccctggggca ttgctgtggg actgctctgc cagtttgggc 480
tcatgccttt tacagcttat ctcctggcca ttagcttttc tctgaagcca gtccaagcta 540
ttgctgttct catcatgggc tgctgcccgg ggggcaccat ctctaacatt ttcaccttct 600
gggttgatgg agatatggat ctcagcatca gtatgacaac ctgttccacc gtggccgccc 660
tgggaatgat gccactctgc atttatctct acacctggtc ctggagtctt cagcagaatc 720
tcaccattcc ttatcagaac ataggaatta cccttgtgtg cctgaccatt cctgtggcct 780
ttggtgtcta tgtgaattac agatggccaa aacaatccaa aatcattctc aagattgggg 840
ccgttgttgg tggggtcctc cttctggtgg tcgcagttgc tggtgtggtc ctggcgaaag 900
gatcttggaa ttcagacatc acccttctga ccatcagttt catctttcct ttgattggcc 960
atgtcacggg ttttctgctg gcacttttta cccaccagtc ttggcaaagg tgcaggacaa 1020

tttccttaga aactggagct cagaatattc agatgtgcat caccatgctc cagttatctt 1080
tcactgctga gcacttggtc cagatgttga gtttcccact ggcctatgga ctctttccagc 1140
tgatagatgg atttcttatt gttgcagcat atcagacgta caagaggaga ttgaagaaca 1200
aacatggaaa aaagaactca ggttgcacag aagtctgcca tacgaggaaa tcgacttctt 1260
ccagagagac caatgccttc ttggaggtga atgaagaagg tgccatcact cctgggccac 1320
cagggccaat ggattgccac agggctctcg agccagttgg ccacatcact tcatgtgaat 1380
agcagggact agctggctgg actggccccc ttctttttca gtggccagta aagacagtgt 1440
gcagctgaca catgaatctt gttggtaggg ccagtgtgaa tatttaagtg ttcaatgtta 1500
gaatatttat attttca 1517

<210> 27
<211> 2142
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7612560CB1

<400> 27
ggtgtacatc tacactagac accttcctgc ttccctcctt ccagagcaga cctctttgtc 60
accccgagct ccttgtttct taagcagtca tgtctgtgac aaaaagtact gagggtcccc 120
agggagccgt tgccatcaaa ttggaccttа tgtcgcctcc tgaaagtgcc aagaagttgg 180
agaacaagga ctctacattc ttggatgaaa gtccttcaga gtcagcaggc ttgaagaaga 240
ccaagggcat aacagtgttc caggccttga ttcacctggt gaaaggcaac atgggcacag 300
ggatcctggg actacccctc gctgtgaaga acgcgggcat cctgatgggc ccactcagtc 360
tgctggtgat gggcttcatt gcctgccact gtatgcacat cctggtcaag tgtgcccagc 420
gcttctgtaa gaggcttaac aagcccttta tggactatgg ggacacggtg atgcatggac 480
tagaagccaa ccccaacgcc tggctccaga atcacgctca ctggggaagg catatcgtga 540
gcttcttcct tattatcacc caacttggct tctgctgtgt gtacattgtg tttttggctg 600
ataatttaaa acaggtagtg gaagctgtta atagcacaac caacaactgc tattccaatg 660
agacggtgat tctgaccccc accatggact cgcgactcta catgctctcc ttcctgccct 720
tcctggtgct gctggtcctc atccggaacc tcaggatctt gaccatcttc tccatgctgg 780
ccaacatcag catgctggtc agcttggtca tcatcataca gtacattacc caggaaatcc 840
cagaccccag ccggttgcca ctggtagcaa gctggaagac ctaccctctc ttcttcggaa 900
cagccatttt ttcttttgaa agcattggtg tggttctgcc tctggaaaac aagatgaaga 960
atgcccgcca cttcccagcc atcctgtctt tgggaatgtc catcgtcact tccctataca 1020
ttggcatggc ggctctgggc tacctgcggt ttggagatga catcaaggcc agcataagcc 1080
ttaacctgcc taactgctgg ctgtaccagt ctgtcaagct tctctacatt gccggcatcc 1140
tgtgcaccta tgccctgcag ttctacgtcc ctgcagaaat catcatcccc tttgccatct 1200
cccgggtgtc aacacgctgg gcactgcctc tggatctgtc cattcgcctc gtcatggtct 1260
gcctgacatg cctcctggcc atcctcatcc cccgcctgga cctggtcatc tccctggtgg 1320
gctccgtgag tggcaccgcc ctggccctca tcatcccacc gctcctggag gtcaccacgt 1380
tctactcaga gggcatgagc cccctcacca tcttcaagga cgtcctgatc agcatcctgg 1440
gcttcgtggg ctttgtggtg gggacctacc aggccctgga cgagctgctc aagtcagaag 1500
actctcaccc cttttccaac tccaccactt ttgttcgcgt ggagctatgc aagaagcagc 1560
caccagaggg ccccaagtgg cagcaactgg ccaaaggaga tgcagccagc taagactgtc 1620
cacactttgg cagacaaccg gttttccctt ttctgggtct gttcaaaaag caaacattaa 1680
gggtgggcac ataatccaca agccagaaag ttgtgcacgg ctccagtgtt gagatgggta 1740
gggccaagat gaccagtgtg aaaactctca gatagaaagg agccatgcat attaaatgag 1800
gggcaacaaa catttcaaac gattagataa cattttctcc caactcaaag atcccaacaa 1860
tgaataggag gcatggaagt agatgtgcca atggggaggg atgaggagtg aacatgaata 1920
ttatttgaat agactttacc tcttaattct tgcaacatgc attcttgatt acctactgtg 1980
tgccaaacaa gatttgtgag aatattgcaa aaatgaccat aaattcctcg tgataatgtg 2040
actttgcacc tgctcctatg aaaagatgaa gtctgtatct gtatccctta aatttttttg 2100
cttgtttgtc ggttttgttt tgtgttttgt ttttttgaga tg 2142

<210> 28
<211> 1661
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2880370CB1

<400> 28
gacactaagc tttaaattca agtaaatagg aggctttttt ttttcgcat aagcagaaat 60
gaggaaatca agaggaagag attagatttc tgttgtgata aatcgaatct gttaaatgcc 120
atgacttttt aattgtctta atcacaagtt aaaccggttg tgttgctgct tagatggcta 180
tatatttgtt taaaagtaca gcagtccctc ctactggact ttgatcctac aaaaacaact 240
gttatctaac tcaccctcag actgtcactg gaacacctgc atgaagaatg ttctttcatt 300
ttttaaaaac gattttgcat atatgattta tttcagcttt caaaatgatt agaaaacttt 360
ttattgttct acttttgttg cttgtgacta tagaagaagc aaggatgtca tcgctcagtt 420
ttctgaatat agagaagact gaaatactat ttttcacaaa gactgaagaa accatccttg 480
taagttcaag ctacgaaaat aaacggccta attccagcca cctctttgtg aaaatagaag 540
atcctaaaat actacaaatg gtgaatgtgg ccaagaagat ctcatcagat gctacaaact 600
ttaccataaa tctggtgact gatgaagaag gagaaacaaa tgtgactatt caactctggg 660
attctgaagg taggcaagaa agactcattg aagaaatcaa gaatgtgaaa gtcaaagtgc 720
tcaaacaaaa agacagtcta ctccaggcac caatgcatat tgatagaaat atcctaatgc 780
ttattttacc actaatacta ttgaataagt gtgcatttgg ttgtaagatt gaattacagc 840
tgtttcaaac agtatggaag agacctttgc cagtaattct tggggcagtt acacagtttt 900
ttctgatgcc attttgcggg tttcttttgt ctcagattgt ggcattgcct gaggcgcaag 960
cttttggagt tgtaatgacc tgcacgtgcc caggaggggg tgggggctat ctctttgctc 1020
tgcttctaga tggagatttc acattggcca ttttgatgac ttgcacatca acattattgg 1080
ctctgatcat gatgcctgtc aattcttata tatacagtag gatattaggg ttgtcaggta 1140
cattccatat tcctgtttct aaaattgtgt caacactcct tttcatactt gtgccagtat 1200
caattggaat agtcatcaag catagaatac ctgaaaaagc aagcttctta gagagaataa 1260
ttagacctct gagttttatt ttaatgttcg taggaattta tttgactttc acagtgggat 1320
tagtgttctt aaaaacagat aatctagagg tgattctgtt gggtctctta gttcctgctt 1380
tgggtttgct gtttgggtac tcctttgcta aagtttgtac gctgcctctt cctgtttgta 1440
aaactgttgc tattgaaagt gggatgttaa atagtttctt agctcttgcc gttattcagc 1500
tgtcttttcc acagtccaag gccaatttag cttctgtggc tccttttaca gtagccatgt 1560
gttctggatg tgaaatgtta ctgatcattc tagtttacaa ggctaagaaa agatgtatct 1620
ttttcttaca agataaaagg aaaagaaatt tcctaatcta a 1661

<210> 29
<211> 1501
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6267489CB1

<400> 29
ccagaggaaa ctagtcacaa aaaccctgac tatcacctga tagattgctt gtgctgcctg 60
ataattactc gcacttttcc caggctagtg caaatcttca ggggccgtcc aggactacag 120
agctgtttca ccctaccttg gcttcaatct cttcccccat gctcgaaggt gcggagctgt 180
acttcaacgt ggaccatggc tacctggagg gcctggttcg aggatgcaag gccagcctcc 240
tgacccagca agactatatc aacctggtcc agtgtgagac cctagaagac ctgaaaattc 300
atctccagac tactgattat ggtaactttt tggctaatca cacaaatcct cttactgttt 360
ccaaaattga cactgagatg aggaaaagac tatgtggaga atttgagtat ttccggaatc 420

attccctgga gcccctcagc acatttctca cctatatgac gtgcagttat atgatagaca 480
atgtgattct gctgatgaat ggtgcattgc agaaaaaatc tgtgaaagaa attctgggga 540
agtgccaccc cttgggccgt ttcacagaaa tggaagctgt caacattgca gagacacctt 600
cagatctctt taatgccatt ctgatcgaaa cgccattagc tccattcttc caagactgca 660
tgtctgaaaa tgctctagat gaactgaata ttgaattgct acgcaataaa ctatacaagt 720
cttaccttga ggcattctat aaatttctgta agaatcatgg tgatgtcaca gcagaagtta 780
tgtgtcccat tcttgagttt gaggccgaca gacgtgcttt tatcatcact cttaactcct 840
ttggcactga attgagcaaa gaagaccgag agaccctcta tccaaccttc ggcaaactct 900
atcctgaggg gttgcggctg ttggctcaag cagaagactt tgaccagatg aagaacgtag 960
cggatcatta cggagtatac aaacctttat ttgaagctgt aggtggcagt gggggaaaga 1020
cattggagga cgtgttttac gagcgtgagg tacaaatgaa tgtgctggca ttcaacagac 1080
agttccacta cggtgtgttt tatgcatatg taaagctgaa ggaacaggaa attagaaata 1140
ttgtgtggat agcagaatgt atttcacaga ggcatcgaac taaaatcaac agttacattc 1200
caatttata acccaagtaa ggttctcaaa tgtagaaaat tataaatgtt aaaaggaagt 1260
tattgaagaa aataaaagaa attatgttat attatctaga ctacacaaaa gtaagccaca 1320
ctatatcttc atgagttgca aatccatgga aacacagtaa accagccctg aaacaaagca 1380
tttccttgtt ttcagtggta ttagatcttg tttccacatg tctgtctcat tcttcactgg 1440
gccttacagg ttagttttaa ttaactctat ggtatttttc tattcttgtc tgatcatgtt 1500
a 1501

<210> 30
<211> 5526
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484777CB1

<400> 30
caggctgttt tgtgcaggct gtccctcttc ttcaaaatcg tgcatcccct ccccgaagca 60
gcaggcagtg tgcctccatt cagccacatt tggtatgcat gagcacggct gcagagagag 120
gggaggtggc tgttttaaga aggttcaggg gctcaggcaa ggctacttga ctagtcttcc 180
aagttccagg aagcctctgc cctaatggaa tttgcaggtg tggagatgac catgggatgc 240
cagagccgtg ggggaccgtt tattttctag gcattgctca ggttttcagt ttcttgtttt 300
cctggtggaa tttggaaggg gtcatgaatc aggctgatgc tcctcgaccc ctaaactgga 360
ccatccggaa gctgtgccac gcagcctttc ttccatctgt cagacttctg aaggctcaga 420
aatcctggat agaaagagca ttttataaaa gagaatgtgt ccacatcata cccagcacca 480
aagaccccca taggtgttgc tgtgggcgtc tgataggcca gcatgttggc ctcaccccca 540
gtatctccgt gcttcagaat gagaaaaatg aaagtcgcct ctcccgaaat gacatccagt 600
ctgaaaagtg gtccatcagc aaacacactc aactcagccc tacggatgct tttgggacca 660
ttgagttcca aggaggtggc cattccaaca aagccatgta tgtgcgagta tcttttgata 720
caaaacctga tctcctctta cacctgatga ccaaggaatg gcagttggag cttcccaagc 780
ttctcatctc tgtccatggg ggcctgcaga actttgaact ccagccaaaa ctcaagcaag 840
tctttgggaa agggctcatc aaagcagcta tgacaactgg agcgtggata ttcactggag 900
gggttaacac aggtgttatt cgtcatgttg gcgatgcctt gaaggatcat gcctctaagt 960
ctcgaggaaa gatatgcacc ataggtattg ccccctgggg aattgtggaa aaccaggagg 1020
acctcattgg aagagatgtt gtccggccat accagaccat gtccaatccc atgagcaagc 1080
tcactgttct caacagcatg cattcccact tcattctggc tgacaacggg accactggaa 1140
aatatggagc agaggtgaaa cttcgaagac aactggaaaa gcatatttca ctccagaaga 1200
taaacacaag aatcggtcaa ggtgttcctg tggtggcact catagtggaa ggaggaccca 1260
atgtgatctc gattgttttg gagtaccttc gagacacccc tcccgtgcca gtggttgtct 1320
gtgatgggag tggacgggca tcggacatcc tggcctttgg gcataaatac tcagaagaag 1380
gcggactgat aaatgaatct ttgagggacc agctgttggt gactatacag aagactttca 1440
catacactcg aacccaagct cagcatctgt tcatcatcct catggagtgc atgaagaaga 1500
aggaattgat tacggtattt cggatgggat cagaaggaca ccaggacatt gatttggcta 1560

tcctgacagc tttactcaaa ggagccaatg cctcggcccc agaccaactg agcttagctt 1620
tagcctggaa cagagtcgac atcgctcgca gccagatctt tatttacggg caacagtggc 1680
cggtgggatc tctggagcaa gccatgttgg atgccttagt tctggacaga gtggattttg 1740
tgaaattact catagagaat ggagtaagca tgcaccgttt tctcaccatc tccagactag 1800
aggaattgta caatacgaga catgggccct caaatacatt gtaccacttg gtcagggatg 1860
tcaaaaaggg gaacctgccc ccagactaca gaatcagcct gattgacatc ggcctggtga 1920
tcgagtacct gatgggcggg gcttatcgct gcaactacac gcgcaagcgc ttccggaccc 1980
tctaccacaa cctcttcggc cccaagaggc ccaaagcctt gaaactgctg ggaatggagg 2040
atgatattcc cttgaggcga ggaagaaaga caaccaagaa acgtgaagaa gaggtggaca 2100
ttgacttgga tgatcctgag atcaaccact tccccttccc tttccatgag ctcatggtgt 2160
gggctgttct catgaagcgg cagaagatgg ccctgttctt ctggcagcac ggtgaggagg 2220
ccatggccaa ggccctggtg gcctgcaagc tctgcaaagc catggctcat gaggcctctg 2280
agaacgacat ggttgacgac atttcccagg agctgaatca caattccaga gactttggcc 2340
agctggctgt ggagctcctg gaccagtcct acaagcagga cgaacagctg gccatgaaac 2400
tgctgacgta tgagctgaag aactggagca acgccacgtg cctgcagctt gccgtggctg 2460
ccaaacaccg cgacttcatc gcgcacacgt gcagccagat gctgctcacc gacatgtgga 2520
tgggccggct ccgcatgcgc aagaactcag gcctcaaggt aattctggga attctacttc 2580
ctccttcaat tctcagcttg gagttcaaga acaaagacga catgccctat atgtctcagg 2640
cccaggaaat ccacctccaa gagaaggagg cagaagaacc agagaagccc acaaaggaaa 2700
aagaggaaga ggacatggag ctcacagcaa tgttgggacg aaacaacggg gagtcctcca 2760
ggaagaagga tgaagaggaa gttcagagca agcaccggtt aatcccccte ggcagaaaaa 2820
tctatgaatt ctacaatgca cccatcgtga agttctggtt ctacacactg gcgtatatcg 2880
gatacctgat gctcttcaac tatatcgtgt tagtgaagat ggaacgctgg ccgtccaccc 2940
aggaatggat cgtaatctcc tatattttca ccctgggaat agaaaagatg agagagattc 3000
tgatgtcaga gccagggaag ttgctacaga aagtgaaggt atggctgcag gagtactgga 3060
atgtcacgga cctcatcgcc atccttctgt tttctgtcgg aatgatcctt cgtctccaag 3120
accagccctt caggagtgac gggagggtca tctactgcgt gaacatcatt tactggtata 3180
tccgtctcct agacatcttc ggcgtgaaca agtatttggg cccgtatgta atgatgattg 3240
gaaaaatgat gatagacatg atgtactttg tcatcattat gctggtggtt ctgatgagct 3300
ttggggtcgc caggcaagcc atccttttte ccaatgagga gccatcatgg aaactggcca 3360
agaacatctt ctacatgccc tattggatga tttatgggga agtgtttgcg gaccagatag 3420
accctccctg tggacagaat gagacccgag aggatggtaa aataatccag ctgcctccct 3480
gcaagacagg agcttggatc gtgccggcca tcatggcctg ctacctctta gtggcaaaca 3540
tcttgctggt caacctcctc attgctgtct ttaacaatac atttttgaa gtaaaatcga 3600
tatccaacca agtctggaag tttcagaggt atcagctcat catgactttc catgaaaggc 3660
cagttctgcc cccaccactg atcatcttca gccacatgac catgatattc cagcacctgt 3720
gctgccgatg gaggaaacac gagagcgacc cggatgaaag ggactacggc ctgaaactct 3780
tcataaccga tgatgagctc aagaaagtac atgactttga agagcaatgc atagaagaat 3840
acttcagaga aaaggatgat cggttcaact catctaatga tgagaggata cgggtgactt 3900
cagaaagggt ggagaacatg tctatgcggc tggaggaagt caacgagaga gagcactcca 3960
tgaaggcttc actccagacc gtggacatcc ggctggcgca gctggaagac cttatcgggc 4020
gcatggccac ggccctggag cgcctgacag gtctggagcg ggccgagtcc aacaaaatcc 4080
gctcgaggac ctcgtcagac tgcacggacg ccgcctacat tgtccgtcag agcagcttca 4140
acagccagga agggaacacc ttcaagctcc aagagagtat agaccctgca ggtgaggaga 4200
ccatgtcccc aacttctcca accttaatgc cccgtatgcg aagccattct ttctattcag 4260
tcaatatgaa agacaaaggt ggtatagaaa agttggaaag tatttttaaa gaaaggtccc 4320
tgagcctaca ccgggctact agttcccact ctgtagcaaa agaacccaaa gctcctgcag 4380
cccctgccaa caccttggcc attgttcctg attccagaag accatcatcg tgtatagaca 4440
tctatgtctc tgctatggat gagctccact gtgatataga ccctctggac aattccgtga 4500
acatccttgg gctaggcgag ccaagctttt caactccagt accttccaca gccccttcaa 4560
gtagtgccta tgcaacactt gcacccacag acagacctcc aagccggagc attgattttg 4620
aggacatcac ctccatggac actagatctt tttcttcaga ctacacccac ctcccagaat 4680
gccaaaaccc ctgggactca gagcctccga tgtaccacac cattgagcgt tccaaaagta 4740
gccgctacct agccaccaca ccctttcttc tagaagaggc tcccattgtg aaatctcata 4800
gctttatgtt ttccccctca aggagctatt atgccaactt tggggtgcct gtaaaaacag 4860
cagaatacac aagtattaca gactgtattg acacaaggtg tgtcaatgcc cctcaagcaa 4920

ttgcggacag agctgccttc cctggaggtc ttggagacaa agtggaggac ttaacttgct 4980
gccatccaga gcgagaagca gaactgagtc accccagctc tgacagtgag gagaatgagg 5040
ccaaaggccg cagagccacc attgcaatat cctcccagga gggtgataac tcagagagaa 5100
ccctgtccaa caacatcact gttcccaaga tagagcgcgc caacagctac tcggcagagg 5160
agccaagtgc gccatatgca cacaccagga agagcttctc catcagtgac aaactcgaca 5220
ggcagcggaa cacagcaagc ctgcgaaatc ccttccagag aagcaagtcc tccaagccgg 5280
agggccgagg ggacagcctg tccatgagga aactgtccag aacatcggct ttccaaagct 5340
ttgaaagcaa gcacacctaa accttcttaa tatccgccac agaaggctca agaatccagc 5400
cctaaaattc tctccaactc cagtttttcc cctttccttg aatcatacct gctttattct 5460
tagctgagca aaacaagcaa tgctttggga ggtgttaact caaaggtgac ttctgggcca 5520
cagatc 5526

<210> 31
<211> 2739
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2493969CB1

<400> 31
gcgcagtaag tgcggactgc cagccaccag ccttggcagc cagctcgtcg cctccagccc 60
cgaccccgac attcatgccc aggagaaggc tgcactgggt ccctctgggc ctttcctaaa 120
agggagatcc ctgttcacta gatgagttcc agaaccatcc actaaggctt tgtagccccc 180
ttccatcagc tgaccttcac tgcatcccct atcgctcaag atgagtggct tcttcacctc 240
gctggacccc cggcgggtgc agtggggagc tgcctggtat gcaatgcact ccaggatcct 300
acgcaccaaa ccagtggagt ccatgctaga gggaactggg accaccacgg cacatggaac 360
taagctagcc caggtactca ccacagtgga cctcatctct cttggcgttg gcagctgtgt 420
gggcactggc atgtatgtgg tctctggcct ggtggccaag gaaatggcag gacctggtgt 480
cattgtgtcc ttcatcattg cagccgtcgc atccatatta tcaggcgtct gctatgcaga 540
gtttggagtt cgagtcccca agaccacagg atctgcctac acctacagct atgtcactgt 600
tggggaattt gtggcatttt tcattggctg gaacctgatc ctggagtacc tgattggcac 660
tgcggccgga gccagtgctc tgagcagcat gtttgactca ctagccaacc acaccatcag 720
ccgctggatg gcggacagcg tgggaaccct caatggcctg gggaaaggtg aagaatcata 780
cccagacctt ctggctctgt tgatcgcggt catcgtgacc atcattgttg ctctgggggt 840
gaagaattcc ataggcttca acaatgttct caatgtgctg aacctggcag tatgggtgtt 900
catcatgatc gcaggcctct tcttcatcaa tgggaaatac tgggcggagg gccagttctt 960
gccccacggc tggtcagggg tgctgcaagg agcagcaaca tgcttctacg ctttcattgg 1020
ctttgacatc atcgccacca ctggagagga agccaagaat cccaacacgt ccatccctta 1080
tgctatcact gcctccctgg tcatctgcct gacagcatat gtgtctgtga gcgtgatctt 1140
aactctgatg gtgccatatt ataccattga cacggaatcc ccactcatgg agatgtttgt 1200
ggctcatggg ttctatgctg ccaaattcgt agtggccatt gggtcggttg caggactgac 1260
agtcagcttg ctggggtccc tcttcccgat gccgagggtc atttatgcca tggctggtga 1320
cgggctcctt ttcaggttcc tggctcacgt cagctcctac acagagacac cagtggtggc 1380
ctgcatcgtg tcggggttcc tggcagcgct cctcgcactg ttggtcagct tgagagacct 1440
gatagagatg atgtctatcg gcacgctcct ggcctacacc ttggtctctg tctgtgtctt 1500
gctccttcga taccaacctg agagtgacat tgatggtttt gtcaagttct tgtctgagga 1560
gcacaccaag aagaaggagg gcattctggc tgactgtgag aaggaagctt gttctcctgt 1620
gagtgagggg gatgagtttt ctggcccagc caccaacaca tgtggggcca agaacttacc 1680
atccttggga gacaatgaga tgctcatagg gaaatcagac aagtcaacct acaacgtcaa 1740
ccaccccaat tacggcaccg tggacatgac cacaggcata gaagctgatg aatccgaaaa 1800
tatttatctc atcaagttaa agaagctgat tgggcctcat tattacacca tgagaatccg 1860
gctgggcctt ccaggcaaaa tggaccggcc cacagcagcg acggggcaca cggtgaccat 1920
ctgcgtgctc ctgctcttca tcctcatgtt catctttctgc tccttcatca tctttggttc 1980
tgactacatc tcagagcaga gctggtgggc catccttctg gttgttctga tggtgctgct 2040

gatcagcacc ctggtgtttg tgatcctgca gcagccagag aaccccaaga agctgcccta 2100
catggccccc tgcctccccc ttgtgcctgc ctttgccatg ctggtgaaca tctatctcat 2160
gctaaagctc tccaccatca catggatccg gtttgcggtc tggtgctttg tgggtctgct 2220
catttatttt ggatatggca tctggaacag caccctggaa atcagcgctc gagaagaggc 2280
cctgcaccaa agcacgtacc aacgctacga cgtggatgac cccttctcag tggaggaggg 2340
tttctcctac gccacagagg gcgagagcca ggaggactgg ggcgggccca ctgaagacaa 2400
aggcttctat taccaacaga tgtcagatgc gaaggcaaac ggccggacaa gtagcaaagc 2460
gaagagcaaa agcaaacaca aacagaactc agaggccctg attgcaaatg atgagttaga 2520
ttactctcca gagtaggaga aacacacaag tgggtagaaa tggtgatgac tgattttcag 2580
taacttaacc tgtgggctag aaggtgaaaa ctttttggc tctcatttca caaatccagc 2640
cttccccaaa ttcaatccct agtcatagcc tgtcatttgc tacttttgct cttcaggata 2700
gttctgttga agggcttaac ctgggtcccc taactggtc 2739

<210> 32
<211> 4321
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3244593CB1

<400> 32
atggtgggtg aaggacccta ccttatctca gatctggacc agcgaggccg gcggagatcc 60
tttgcagaaa gatatgaccc cagcctgaag accatgatcc cagtgcgacc ctgtgcaagg 120
ttagcaccca acccggtgga tgatgccggg ctactctcct tcgccacatt ttcctggctc 180
acgccggtga tggtgaaagg ctaccggcaa aggctgaccg tagacaccct gcccccattg 240
tcgacatatg actcatctga caccaatgcc aaaagatttc gagtcctttg ggatgaagag 300
gtagcaaggg tgggtcctga gaaggcctct ctgagccacg tggtgtggaa attccagagg 360
acacgcgtgt tgatggacat cgtggccaac atcctgtgca tcatcatggc agccataggg 420
ccgacagttc tcattcacca aatcctccag cagactgaga ggacctctgg gaaagtctgg 480
gttggcattg gactgtgcat agccctttt gccaccgagt ttaccaaagt cttcttttgg 540
gcccttgcct gggccatcaa ctaccgcacg gccatccggt tgaaggtggc gctctccacc 600
ttggtttttg aaaacctagt gtccttcaag acattgaccc acatctctgt tggcgaggtg 660
ctcaatatac tgtcaagtga tagctattct ttgtttgaag ctgccttgtt ttgtcctttg 720
ccagccacca tcccgatcct aatggtcttt tgtgcggcgt acgccttttt cattctgggg 780
cccacagctc tcatcgggat atcagtgtat gtcatattca tacccgtcca gatgtttatg 840
gccaagctca attcagcttt ccgaaggtca gcaattttgg tgacagacaa gcgagttcag 900
acaatgaatg agtttctgac ctgcatcagg ctgatcaaaa tgtatgcctg ggagaaatct 960
tttaccaaca ctatccaaga tataagaagg agggaaagaa aattactgga aaaagctgga 1020
tttgtccaaa gtggaaactc tgccctggcc cccatcgtgt ccaccatagc catcgtgctg 1080
acattatcct gccacatcct cctgagacgc aaactcaccg cacccgtggc atttagtgtg 1140
attgccatgt ttaatgtaat gaagttttcc attgcaatct tgcccttctc catcaaagca 1200
atggctgaag cgaatgtctc tctaaggaga atgaagaaaa ttctcataga taaaagcccc 1260
ccatcttaca tcacccaacc agaagaccca gatactgtct tgcttttagc aaatgccacc 1320
ttgacatggg agcatgaagc cagcaggaaa agtaccccaa agaaattgca gaaccagaaa 1380
aggcatttat gcaagaaaca gaggtcagag gcatacagtg agaggagtcc accagccaag 1440
ggagccactg gcccagagga gcaaagtgac agcctcaaat cggttctgca cagcataagc 1500
tttgtggtga gaaaggggaa gatcttggga atatgtggga atgtgggaag tggaaagagc 1560
tccctccttg cagctctcct aggacagatg cagctgcaga aaggggtggt ggcagtcaat 1620
ggaactttgg cctacgtttc acagcaggca tggatctttc atggaaatgt gagagaaaac 1680
atactctttg gagaaaagta tgatcaccaa aggtatcagc acacagtccg cgtctgtggc 1740
ctccagaagg acctgagcaa cctcccctat ggagacctga ctgagattgg ggagcggggc 1800
ctcaacctct ctgggggaca gaggcagagg attagcctgg cccgcgctgt ctactccgac 1860
cgtcagctct acctgctgga cgacccccctg tcggccgtgg acgcccacgt ggggaagcac 1920
gtctttgagg agtgcattaa gaagacgctc aggggaaaga cagtcgtcct ggtgacccac 1980

cagctacagt tcttagagtc ttgtgatgaa gttattttat tagaagatgg agagatttgt 2040
gaaaagggaa cccacaagga gttaatggag gagagagggc gctatgcaaa actgattcac 2100
aacctgcgag gattgcagtt caaggatcct gaacaccttt acaatgcagc aatggtggaa 2160
gccttcaagg agagccctgc tgagagagag gaagatgctg gtataatcgt tttggctcca 2220
ggaaatgaga aagatgaagg aaaagaatct gaaacaggct cagaatttgt agacacaaaa 2280
gggtacctcc tttctctctt cactgtgttc ctcttcctcc tgatgattgg cagcgctgcc 2340
ttcagcaact ggtggctggg tctctggttg gacaagggct cacggatgac ctgtgggccc 2400
cagggcaaca ggaccatgtg tgaggtcggc gcggtgctgg cagacatcgg tcagcatgtg 2460
taccagtggg tgtacactgc aagcatggtg ttcatgctgg tgtttggcgt caccaaaggc 2520
ttcgtcttca ccaagaccac actgatggca tcctcctctc tgcatgacac ggtgtttgat 2580
aagatcttaa agagcccaat gagtttcttt gacacgactc ccactggcag gctaatgaac 2640
cgtttttcca aggatatgga cgagctggat gtgaggctgc cgtttcacgc agagaacttt 2700
ctgcagcagt tttttatggt ggtgtttatt ctcgtgatct tggctgctgt gtttcctgct 2760
gtccttttag tcgtggccag ccttgctgta ggcttcttca ttctgttacg cattttccac 2820
agaggagtcc aggagctcaa gaaggtggag aatgtcagcc ggtcaccctg gttcacccac 2880
atcacctcct ccatgcaggg cctgggcatc attcacgcct atggcaagaa ggagagctgc 2940
atcacctatc acctcctcta ctttaactgt gctctcaggt ggtttgcgct gagaatggat 3000
gtcctcatga acatccttac cttcactgtg gccttgttgg tgaccctgag tttctcctcc 3060
atcagtactt catccaaagg cctgtcattg tcatacatca tccagctgag cggactgctc 3120
caagtgtgtg tgcgaacggg aacagagacg caagccaaat tcacctccgt ggagctgctc 3180
agggaataca tttcgacctg tgttcctgaa tgcactcatc ccctcaaagt ggggacctgt 3240
cccaaggact ggcccagctg tggggagatc accttcagag actatcagat gagatacaga 3300
gacaacaccc cccttgttct cgacagcctg aacttgaaca tacaaagtgg gcagacagtc 3360
gggattgttg gaagaacagg ttccggaaag tcatcgttag gaatggcttt gtttcgtctg 3420
gtggagccag ccagtggcac aatctttatt gatgaggtgg atatctgcat tctcagcttg 3480
gaagacctca gaaccaagct gactgtgatc ccacaggatc ctgtcctgtt tgtaggtaca 3540
gtaaggtaca acttggatcc ctttgagagt cacaccgatg agatgctctg gcaggttctg 3600
gagagaacat tcatgagaga cacaataatg aaactcccag aaaaattaca ggcagaagtc 3660
acagaaaatg gagaaaactt ctcagtaggg gaacgtcagc tgctttgtgt ggcccgagct 3720
cttctccgta attcaaagat cattctcctt gatgaagcca ccgcctctat ggactccaag 3780
actgacaccc tggttcagaa caccatcaaa gatgccttca agggctgcac tgtgctgacc 3840
atcgcccacc gcctcaacac agttctcaac tgcgatcacg tcctggttat ggaaaatggg 3900
aaggtgattg agtttgacaa gcctgaagtc cttgcagaga agccagattc tgcatttgcg 3960
atgttactag cagcagaagt cagattgtag aggtcctggc ggctgattct agaggaggaa 4020
gaggctctgt gagatgaata ggaggagtct tcaggaggag gggctgtcct ctccgcaggc 4080
agccctggtc ttcagcccct cccatccacg gagtgagctg gggctgaagt tgtccccact 4140
gccatactca gtccatgtca ccccacttgg tgggcttggg gttggttctg ggtggtgaac 4200
cggggcagac ccagctaatg gattaaaaaa ctgccttca cctcccaaat ccccaagggt 4260
tcctcatgtg ttttcaccaa aaccacccca gtgcctgaga ttgaaaatat tgtaactttc 4320
a 4321

<210> 33
<211> 4519
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4921451CB1

<400> 33
ttcaggaccg ttggcaccgg gctaacggtt ccaccacgtc cgccgccctg gacgcccgcg 60
gcctgccccc ccctgcctct cctgcgccga tacacttcga gtggattctg gccatttgag 120
cattctctcc aactctccaa tccccagtct gcccccacgg gggtctcccc cacctctccc 180
ccgtcccaca gcctaaaccc ctcttcgccc tgaacctccc ttttcctcat gcggtgaatg 240
ggcactggcc ccgctcagac tcccaggagc accagagctg gccctgagcc aagccctgcc 300

ccaccaggac ctggggacac gggtgactca gacgtgactc aggaaggctc aggtcctgct 360
ggcatccgcg gagccccacc agcatgggca gcctcggcca gagagaagat ctccgagatg 420
aggacaggaa ctcaggtgct gatcctgggc ggaggggcgg gtgcagcatt cacctggaag 480
gtccaggcca acaaccgtgc ctacaacggg cagttcaagg agaaggtgat cctgtgctgg 540
caaaggaaga aatacaagac caatgtcatc cgcacggcca agtacaactt ctactcgttc 600
ctgccgctga acctgtacga gcagttccac cgcgtgtcca acctgttctt cctcatcatc 660
atcatcctgc agagcattcc cgacatctcc acgctgccct ggttctcgct cagtacccct 720
atggtctgcc tcctcttcat ccgtgccacc cgggacctgg tggacgacat ggggagacac 780
aagagtgaca gagccatcaa caacagaccc tgccagattc tgatggggaa gagcttcaag 840
cagaagaaat ggcaggatct gtgcgtgggg gatgtggtct gtctccgcaa ggacaacatc 900
gtcccagtga gctggggtgg accccgaggt cccagaacca cgcgccccct caccgagagc 960
acccctccca gggtggggag ggctgccgca cccccaattt gtcttgcatc ccctcttgca 1020
acgctgcccc ccactccaca ccaggccgac atgctcttgc tggccagcac ggagcccagc 1080
agcctgtgct atgtggagac ggtggacatt gacggggaga ccaacttgaa gttcagacag 1140
gccctgatgg tcacccacaa agaactggcc actataaaga agatggcgtc ctttcaaggc 1200
acagtgacgt gtgaggcgcc taacagtcgg atgcaccact tcgtggggtg cctggaatgg 1260
aatgacaaga aatactccct ggacattggc aacctcctcc tccgaggctg caggattcgc 1320
aacacagaca cctgctatgg actggtcatt tatgctggtt ttgacacaaa aattatgaag 1380
aactgtggca agatccattt gaagagaacc aagctggacc tcctgatgaa caagctggtg 1440
gttgtgatct tcatctccgt ggtgcttgtc tgcctggtgt tggccttcgg cttcggtttc 1500
tcagtcaaag aattcaaaga ccaccactac tacctctcgg gggtgcatgg gagcagcgtg 1560
gccgcagagt ccttcttcgt cttctggagc ttcctcatcc tgctcagcgt caccatcccg 1620
atgtccatgt tcatcctgtc cgagttcatc tacctgggga acagcgtctt catcgactgg 1680
gacgtgcaga tgtactacaa gccgcaggac gtgcctgcca aggcccgcag caccagcctc 1740
aacgaccacc tgggccaggt ggaatacatc ttctcggaca agacgggcac gctcacgcag 1800
aacatcttga ccttcaacaa gtgctgcatc agcggccgcg tctatggaga acccctacct 1860
ctggaacaag ttcgccgacg ggaagctgct cttccacaat gcggccctgc tgcacctcgt 1920
gcggaccaac ggggacgagg ccgtgcggga gttctggcgc ctgctggcca tctgccacac 1980
ggtgatgacc agctgttgta ccaggcggcc tcccccgacg agggggcgct ggtcaccgca 2040
gcccggaact tcggctacgt gttcctgtcc cgcacccagg acaccgtcac gatcatggag 2100
ctgggggagg aacgggtcta ccaggtcctg gccataatgg acttcaacag cacgcgcaaa 2160
cggatgtcgg tgctggttcg aaagccagag ggcgccatct gcctgtacac caagggcgcc 2220
gacacggtca tcttcgaacg cttgcacagg agggggggcaa tggaatttgc cacagaggag 2280
gccttggctg cctttgccca ggagaccctg cggacactgt gcctggccta cagggaggtg 2340
gctgaggaca tttacgagga ctggcagcag cgccaccagg aggccagcct cctgctgcag 2400
aaccgggcac aggccctgca acaggtgtac aacgagatgg agcaggacct caggctgctg 2460
ggagccacag ccatcgagga cagactccag gacggtgtcc ctgaaaccat caaatgtctc 2520
aagaagagca acatcaaaat atgggtgctc accggggaca agcaggaaac ggctgtgaac 2580
atcggcttcg cctgcgagct gctgtcagag aatatgctca ttctggagga gaaggagatt 2640
agccgcatcc tggagaccta ctgggaaaac agtaacaacc ttctaaccag ggagtccctg 2700
tcgcaggtca agctggcctt ggtcattaac ggagacttcc tggacaaact gctggtgtcc 2760
ctgcggaagg agccgcgcgc cctggcgcag aacgtgaaca tggacgaggc gtggcaggag 2820
ctcggccagt ccaggaggga tttcctctac gccaggcgcc tgtccctgct gtgccggagg 2880
ttcgggctcc cgctggctgc accgccagcc caggactcca gagcccgccg tagctccgag 2940
gtgctgcagg agcgcgcctt cgtggacctg gcgtccaagt gccaggcggt catctgctgc 3000
cgcgtgacgc ccaagcagaa ggccctgatc gtggccctgg tcaagaagta ccaccaggtg 3060
gtgaccctgg ccatcgggga cggtgccaac gacatcaaca tgatcaagac cgcggacgtg 3120
ggcgtggggc tggcgggcca ggagggcatg caggcagttc agaacagcga cttcgtgctc 3180
ggccagttct gcttcctgca gcgcctcctg ctggtgcacg gccgctggtc ctacgtgcgg 3240
atctgcaagt tcctgcgcta cttcttctac aagagcatgg ccagcatgat ggtgcaggtc 3300
tggtttgcct gctacaacgg cttcaccggc caggacgtga gcgcagagca gagcctggag 3360
aagccggagc tgtacgtggt ggggcagaag gacgagctct tcaactactg ggtcttcgtc 3420
caagccatcg cccatggtgt gaccacctct ctggtcaact tcttcatgac actgtggatc 3480
agccgcgaca cggcgggacc cgccagcttc agcgaccacc agtcctttgc ggtcgtggtg 3540
gccctgtctt gcctgctgtc catcaccatg gaggtcattc ttatcatcaa gtactggacc 3600
gccctgtgcg tggcgaccat cctcctcagc cttggtttct acgccatcat gactaccacc 3660

acccagagct tctggctctt cagagtatcc cccacgacct tcccgtttct gtatgccgac 3720
ctcagcgtga tgtcctctcc ctccatcctg ctggtggtcc tgctgagtgt gtccataaac 3780
accttccctg tcctggccct ccgagtcatc ttcccagccc tcaaggagct acgtgccaag 3840
gaggagaagg tggaggaggg ccccagcgag gagattttca ccatggagcc cttgcctcat 3900
gtacaccggg agtctcgtgc ccgccgttcc agctatgctt tctcccaccg ccagctgacg 3960
ttggagagcc agccagactc ctcggaggag aagtcagcat ttttgaagcc ctccacaccg 4020
ttccggaaga gctggcaaaa ggagcctcac acccccaagg aggggacggt gccacttcca 4080
gacaagaccc acaaatctca ggtggagact ctgccaccaa gtctggaaga atcgtccacg 4140
tccacgagcg agcagcctat ggaggtggag ctgtggcccg cggagaagca gtcatcatca 4200
tccatggagt ggctgctggt gcccggggag gagcagctat ccttgccccc agaggagcag 4260
tcattgccct ctgcggaggg gaccagggtt cagcagtgac gtagcatctg aatccctaga 4320
cccatctgat gaagaggcat cttcgagccc aaaggagtca cgctggcata tcaggaagat 4380
gtccttcctg ggaagaagaa gctccagcca gttctgctgc aagtcaacca gcatgcaggg 4440
ggccttcctc taaagacaag gactccacat gcttttcttt ttctaataaa ccagggtcca 4500
tctgacccca gcgctaaaa 4519

<210> 34
<211> 2922
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5547443CB1

<400> 34
gaggagtctg gcatggctca tgaatcagca gaggacttgt ttcatttcaa cgtagggggc 60
tggcatttct cagttcccag aagcaaactc tctcagtttc cagactccct gctgtggaaa 120
gaggcttcag ccttgacctc ttcagaaagc cagaggctat ttatcgacag agatggttcc 180
acatttaggc acgtgcacta ttacctctac acctccaaac tctccttctc cagttgtgca 240
gaactgaact tgctgtatga gcaagcattg ggtttgcagc tgatgccttt gctgcagact 300
ctagataacc tgaaggaagg gaaacaccat ctacgcgtac ggcctgcaga cctacctgtt 360
gctgagagag catctctgaa ctactggcgt acatggaagt gtattagcaa accctcagaa 420
tttccaatta aaagcccagc ctttacaggc ctacatgata aggcacctct ggggctcatg 480
gacacacccc tgttagacac agaagaggag gtgcactact gcttcctgcc cctagacctg 540
gtggccaaat atcccagcct agtgactgaa gacaacctgc tgtggctggc tgagacggtg 600
gccctcatcg agtgcgagtg cagcgagttc cgcttcattg tgaattttct tcgctcacag 660
aagattttac taccggataa tttctccaac attgatgtat tagaagcaga agtggaaatt 720
ctggaaatcc ctgcactcac tgaagccgta aggtggtacc ggatgaacat gggtggctgt 780
tccccgacca cctgttctcc cctgagcccc gggaaggggg cccgcacagc cagcctggag 840
tccgtgaaac cgctctacac aatggccctg ggtctgctgg tcaagtaccc ggactctgcg 900
ctgggccagc ttcgcatcga gagcacgcta gacggaagcc gactgtacat cacagggaat 960
ggcgtcctct ttcagcacgt caagaactgg ctggggactt gccggctgcc cctgacagag 1020
accatttccg aggtatatga gctctgtgcc ttcctagaca aaagggacat cacctacgag 1080
ccaatcaaag ttgctttgaa gactcatctg gagccaagga ctttggcacc catggatgtg 1140
ctcaatgagt ggacggcaga gatcactgtg tattccccac aacagatcat caaagtgtat 1200
gttggaagcc actggtacgc aaccaccctg cagacactgc tgaagtatcc agaactgctg 1260
tccaaccctc agagagtgta ctggatcaca tatggacaaa ccctgctcat ccacggggat 1320
ggccagatgt tccgacacat tctcaacttc ctgagacttg gcaaactgtt tttaccatct 1380
gaatttaagg aatggcccct cttctgccag gaggtggagg aataccacat tccatccctc 1440
tcagaagccc ttgcacaatg tgaagcatac aagtcatgga ctcaggagaa agaatctgaa 1500
aatgaagaag ctttttccat caggaggctg catgtggtga cagaagggcc agggtcactg 1560
gtggagttca gtagagacac taaagaaacc acagcctaca tgcctgtgga cttcgaagac 1620
tgcagtgaca ggactccatg gaacaaggct aagggaaacc tggtcaggtc caaccagatg 1680
gatgaggctg agcagtacac tcggcccatc caggtgtccc tatgccgaaa tgccaagagg 1740
gctggcaacc ctagcacata ctcacactgc cgtggcttgt gtaccaatcc tggacactgg 1800

gggagccacc ctgagagccc cccaaagaag aaatgcacca caatcaacct cacacagaaa 1860
tctgaaacca aagaccctcc cgccactccc atgcaaaaac tcatctccct ggtgagagaa 1920
tgggacatgg tcaattgcaa acagtgggaa ttccagccac tgacagccac acggagcagc 1980
cccttggagg aggccaccct gcagctcccc ttgggaagcg aggctgcttc ccagcccagc 2040
acctcagctg cctggaaagc ccattccaca gcctcagaga aggatccagg accacaggca 2100
ggggctggag ctggagcgaa agacaagggg ccagagccaa ccttcaagcc atacttaccc 2160
ccaaaaagag ctggcaccct gaaggactgg agcaagcaga ggaccaagga gagagaaagc 2220
cctgcccctg agcagcctct gcccgaggcc agtgaggtgg acagcctagg ggttatcctc 2280
aaagtgactc acccccccgt ggtgggcagc gatggcttct gcatgtttct tgaggacagc 2340
atcatctata ccacggagat ggacaacctc aggcacacaa cacccacagc cagtccccag 2400
ccccaagaag tgactttcct gagtttctct ctgtcctggg aagagatgtt ttatgcacag 2460
aaatgtcact gcttcctggc tgacatcatc atggattcca tcaggcaaaa ggaccccaaa 2520
gccatcacag ccaaggtggt ctccctggcc aatcggctgt ggaccctgca catcagcccc 2580
aagcagtttg tggtagattt gctggccatc accggcttca aggatgaccg gcacacccag 2640
gagcgcctgt acagctgggt ggagcttaca ctgcccttcg ccaggaaata tggccgatgc 2700
atggacctgc tcatccagag gggcctgtct aggtctgtct cttactccat cctgggaaag 2760
tacctacaag aggactaggg tgcccagaga tgcagccccc catgccccac ccgccaagtc 2820
tcatttttaat tggagatagc ccagaatgca tgtgcccatc agagggtaca tatcagtcta 2880
ttttttaata taaacaaata aaagattaaa tcacacatca aa 2922

<210> 35
<211> 2763
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 56008413CB1

<400> 35
ggaccccagg ccgggccggg ccgagaggct gccatgggct ccgtggggag ccagcgcctt 60
gaggagccca gcgtggcagg cacaccagac ccgggcgtag tgatgagctt caccttcgac 120
agtcaccagc tggaggaggc ggcggaggcg gctcagggcc agggccttag ggccaggggc 180
gtcccagctt tcacggatac tacattggac gagccagtgc ccgatgaccg ttatcacgcc 240
atctactttg cgatgctgct ggctggcgtg ggcttcctgc tgccatacaa cagcttcatc 300
acggacgtgg actacctgca tcacaagtac ccagggacct ccatcgtgtt tgacatgagc 360
ctcacctaca tcttggtggc actggcagct gtcctcctga acaacgtcct ggtggagaga 420
ctgaccctgc acaccaggat caccgcaggc tacctcttag ccttgggccc tctccttttt 480
atcagcatct gcgacgtgtg gctgcagctc ttctctcggg accaggccta cgccatcaac 540
ctggccgctg tgggcaccgt ggccttcggc tgcacagtgc agcaatccag cttctacggg 600
tacacgggga tgctgcccaa gcggtacacg caggggggtga tgaccgggga gagcacggcg 660
ggcgtgatga tctctctgag ccgcatcctc acgaagctgc tgctgcccga cgagcgcgcc 720
agcacgctca tcttcttcct ggtgtcggtg gcgctggagc tgctgtgttt cctgctgcac 780
ctgttagtgc ggcgcagccg cttcgtgctc ttctatacca cacggccgcg tgacagccac 840
cggggcaggc caggcctggg caggggctat ggctaccgcg tgcaccacga cgttgtcgcc 900
ggggacgtcc acttcgagca cccagccccg gccctggccc ccaacgagtc cccaaaggac 960
agcccagccc acgaggtgac cggcagcggc ggggcctaca tgcgctttga cgtgccgcgg 1020
ccaagggtcc agcgcagctg gcccaccttc agagccctgt tactgcaccg ctacgtggtg 1080
gcgcgggtga tctgggccga catgctctcc atcgccgtga cctacttcat cacgctgtgc 1140
ctgttccccg gcctcgagtc tgagatccgc cactgcatcc tgggcgagtg gctgcccatc 1200
ctcatcatgg ctgtgttcaa cctgtcagac ttcgtgggca agatcctggc agccctgccc 1260
gtggactggc ggggcaccca cctgctggcc tgctcctgcc tgcgtgtggt cttcatcccc 1320
ctcttcatcc tgtgcgtcta ccccagcggc atgcccgccc tccgtcaccc cgcctggccc 1380
tgcatcttct cactgctcat gggcatcagc aacggctact tcggcagcgt gcccatgatc 1440
ctggcggcag gcaaagtgag ccccaagcag cgggagctgg cagggaacac catgaccgtg 1500
tcctacatgt cagggctgac gctggggtcc gccgtggcct actgcaccta cagcctcacc 1560

cgcgacgctc acggcagctg cctgcacgcc tccaccgcca atggttccat cctcgcaggc 1620
ctctgagcca gccccgccca ctgccaggga cgccgagggc ctgaccaggg gccccgaggc 1680
ctgagggccc ctcccctgtc cccacctcag tgcctgcggg gccctgagcc tccccctgtg 1740
ccagcagccc cactccctca gggtccagcc atgccccacc ctggactgaa gttctgcaaa 1800
gtcctccgag gaccggaaca cgtttctgcg acccggggct ctggccagca ctgtgttctg 1860
cgtttggtct catacctgcg tctaccttcc atctgtgtcc agcggccccg gctccagccc 1920
agccagcact ctgcagggtc acacgcaccg tgtccccacc caggacagca gacacccgcc 1980
agagtgtgcg cgcccagtga ctgcaccccg gcctcatca cccaccggca ctgatcgggg 2040
caccgcctgg cccagcctcc accagggacc cctcctcatg aactctggag ccctgagagg 2100
agaggggcag ccccccacct tgtcaccctc agggcttccc cttctgtcct cattcttaga 2160
gactgcttct cccaaacata acgcgttagc catgaaggag tcggagccct gggtccgaat 2220
ggacccgcct gcggtctgca tcagcctctg ggaaaccaca gcagtgatgc cagctgggca 2280
cgtcaggacc tccccacaca cccacacgat gccacaggtc agggggctgt gcctgactag 2340
ggagccctcc cattgccttc ctggcccggg atagaagagg ggaggtaagt ctgggggcta 2400
cgaagccggg cccccacacc ctggctgaag tcagcttgac ctaggtcttg accctcatcc 2460
agcaagggac tcgacagacc caagggtccc tggaacgtag ggaggggctg ggggtcactc 2520
cagcccgggc ctcccagaac accaggcccg tgtgggtggc accctgaggt cagggatcc 2580
taagggtgtc cttccagaga cggtgtttcc agggggagga ccgccccccgc ttccagatcc 2640
ccggccccgg ctgtgactgc cctgtttcac ccctgctgtg tcccatcccc cgtctgtcca 2700
ctaactgtac cgcaccggcc atttaaagat gaaggcagac cgctgccaaa aaaaaaaaaa 2760
aaa 2763

<210> 36
<211> 5211
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6127911CB1

<400> 36
aagagctgct ggagtaggca cccatttaaa gaaaaaatga agaagcagca ataaagaagt 60
tgtaatcgtt acctagacaa acagagaact ggttttgaca gtgtttctag agtgcttttt 120
attattttcc tgacagttgt gttccaccat gattactttc tccttcagcg aataggctaa 180
atgaatatga aacagaaaag cgtgtatcag caaaccaaag cacttctgtg caagaatttt 240
cttaagaaat ggaggatgaa aagagagagc ttattggaat ggggcctctc aatacttcta 300
ggactgtgta ttgctctgtt ttccagttcc atgagaaatg tccagtttcc tggaatggct 360
cctcagaatc tgggaagggt agataaattt aatagctctt ctttaatggt tgtgtataca 420
ccaatatcta atttaaccca gcagataatg aataaaacag cacttgctcc tcttttgaaa 480
ggaacaagtg tcattggggc accaaataaa acacacatgg acgaaatact tctggaaaat 540
ttaccatatg ctatgggaat catctttaat gaaacttttct cttataagtt aatatttttc 600
cagggatata acagtccact ttggaaagaa gatttctcag ctcattgctg ggatggatat 660
ggtgagtttt catgtacatt gaccaaatac tggaatagag gatttgtggc tttacaaaca 720
gctattaata ctgccattat agaaatcaca accaatcacc ctgtgatgga ggagttgatg 780
tcagttactg ctataactat gaagacatta cctttcataa ctaaaaatct tcttcacaat 840
gagatgttta ttttattctt cttgcttcat ttctccccac ttgtatattt tatatcactc 900
aatgtaacaa aagagagaaa aaagtctaag aatttgatga aaatgatggg tctccaagat 960
tcagcattct ggctctcctg gggtctaatc tatgctggct tcatctttat tatttccata 1020
ttcattacaa ttatcataac attcacccaa attatagtca tgactggctt catggtcata 1080
tttatactct ttttttata tggcttatct ttggtagctt tggtgttcct gatgagtgtg 1140
ctgttaaaga aagctgtcct caccaatttg gttgtgtttc tccttaccct cttttgggga 1200
tgtctgggat tcactgtatt ttatgaacaa cttccttcat ctctggagtg gattttgaat 1260
atttgtagcc cttttgcctt tactactgga atgattcaga ttatcaaact ggattataac 1320
ttgaatggtg taatttttcc tgacccttca ggagactcat acacaatgat agcaacttt 1380
tctatgttgc ttttggatgg tctcatctac ttgctattgg cattatactt tgacaaaatt 1440

ttaccctatg gagatgagcg ccattattct cctttatttt tcttgaattc atcatcttgt 1500
ttccaacacc aaaggactaa tgctaaggtt attgagaaag aaatcgatgc tgagcatccc 1560
tctgatgatt attttgaacc agtagctcct gaattccaag gaaaagaagc catcagaatc 1620
agaaatgtta agaaggaata taaaggaaaa tctggaaaag tggaagcatt gaaaggcttg 1680
ctctttgaca tatatgaagg tcaaatcacg gcaatcctgg gtcacagtgg agctggcaaa 1740
tcttcactgc taaatattct taatggattg tctgttccaa cagaaggatc agttaccatc 1800
tataataaaa atctctctga aatgcaagac ttggaggaaa tcagaaagat aactggcgtc 1860
tgtcctcaat tcaatgttca atttgacata ctcaccgtga aggaaaacct cagcctgttt 1920
gctaaaataa aagggattca tctaaaggaa gtggaacaag aggtacaacg aatattattg 1980
gaattggaca tgcaaaacat tcaagataac cttgctaaac atttaagtga aggacagaaa 2040
agaaagctga cttttgggat taccatttta ggagatcctc aaattttgct tttagatgaa 2100
ccaactactg gattggatcc cttttccaga gatcaagtgt ggagcctcct gagagagcgt 2160
agagcagatc atgtgatcct tttcagtacc cagtccatgg atgaggctga catcctggct 2220
gatagaaaag tgatcatgtc caatgggaga ctgaagtgtg caggttcttc tatgtttttg 2280
aaaagaaggt ggggtcttgg atatcaccta agtttacata ggaatgaaat atgtaaccca 2340
gaacaaataa catccttcat tactcatcac atccccgatg ctaaattaaa aacagaaaac 2400
aaagaaaagc ttgtatatac tttgccactg gaaaggacaa atacatttcc agatcttttc 2460
agtgatctgg ataagtgttc tgaccaggga gtgacaggtt atgacatttc catgtcaact 2520
ctaaatgaag tctttatgaa actggaagga cagtcaacta tcgaacaaga tttcgaacaa 2580
gtggagatga taagagactc agaaagcctc aatgaaatgg agctggctca ctcttccttc 2640
tctgaaatgc agacagctgt gagtgacatg ggcctctgga gaatgcaagt ctttgccatg 2700
gcacggctcc gtttcttaaa gttaaaacgt caaactaaag tgttattgac cctattattg 2760
gtatttggaa tcgcaatatt ccctttgatt gttgaaaata taatatatgc tatgttaaat 2820
gaaaagatcg attgggaatt taaaaacgaa ttgtattttc tctctcctgg acaacttccc 2880
caggaacccc gtaccagcct gttgatcatc aataacacag aatcaaatat tgaagatttt 2940
ataaaatcac tgaagcatca aaatatactt ttggaagtag atgactttga aaacagaaat 3000
ggtactgatg gcctctcata caatggagct atcatagttt ctggtaaaca aaaggattat 3060
agattttcag ttgtgtgtaa taccaagaga ttgcactgtt ttccaattct tatgaatatt 3120
atcagcaatg ggctacttca aatgtttaat cacacacaac atattcgaat tgagtcaagc 3180
ccatttcctc ttagccacat aggactctgg actgggttgc cggatggttc ctttttctta 3240
tttttggttc tatgtagcat ttctccttat atcaccatgg gcagcatcag tgattacaag 3300
aaaaatgcta agtcccagct atggatttca ggcctctaca cttctgctta ctggtgtggg 3360
caggcactag tggacgtcag cttcttcatt ttaattctcc ttttaatgta tttaattttc 3420
tacatagaaa acatgcagta ccttcttatt acaagccaaa ttgtgtttgc tttggttata 3480
gttactcctg gttatgcagc ttcttcttgtc ttcttcatat atatgatatc atttattttt 3540
cgcaaaagga gaaaaaacag tggcctttgg tcattttact tctttttgc ctccaccatc 3600
atgttttcca tcactttaat caatcatttt gacctaagta tattgattac caccatggta 3660
ttggttcctt catatacctt gcttggattt aaaacttttt tggaagtgag agaccaggag 3720
cactacagag aatttccaga ggcaaatttt gaattgagtg ccactgattt tctagtctgc 3780
ttcataccct actttcagac tttgctattc gtttttgttc taagatgcat ggaactaaaa 3840
tgtggaaaga aaagaatgcg aaaagatcct gttttcagaa tttcccccca aagtagagat 3900
gctaagccaa atccagaaga acccatagat gaagatgaag atattcaaac agaaagaata 3960
agaacagcca ctgctctgac cacttcaatc ttagatgaga aacctgttat aattgccagc 4020
tgtctacaca aagaatatgc aggccagaag aaaagttgct tttcaaagag gaagaagaaa 4080
atagcagcaa gaaatatctc tttctgtgtt caagaaggtg aaattttggg attgctagga 4140
cccagtggtg ctggaaaaag ttcatctatt agaatgatat ctgggatcac aaagccaact 4200
gctggagagg tggaactgaa aggctgcagt tcagttttgg gccacctggg gtactgccct 4260
caagagaacg tgctgtggcc catgctgacg ttgagggaac acctggaggt gtatgctgcc 4320
gtcaaggggc tcaggaaagc ggacgcgagg ctcgccatcg caagattagt gagtgctttc 4380
aaactgcatg agcagctgaa tgttcctgtg cagaaattaa cagcaggaat cacgagaaag 4440
ttgtgttttg tgctgagcct cctgggaaac tcacctgtct tgctcctgga tgaaccatct 4500
acgggcatag accccacagg gcagcagcaa atgtggcagg caatccaggc agtcgttaaa 4560
aacacagaga gaggtgtcct cctgaccacc cataacctgg ctgaggcgga agccttgtgt 4620
gaccgtgtgg ccatcatggt gtctggaagg cttagatgca ttggctccat ccaacacctg 4680
aaaaacaaac ttggcaagga ttacattcta gagctaaaag tgaaggaaac gtctcaagtg 4740
actttggtcc acactgagat tctgaagctt ttcccacagg ctgcagggca ggaaaggtat 4800

tcctctttgt taacctataa gctgcccgtg gcagacgttt accctctatc acagaccttt 4860
cacaaattag aagcagtgaa gcataacttt aacctggaag aatacagcct ttctcagtgc 4920
acactggaga aggtattctt agagctttct aaagaacagg aagtaggaaa ttttgatgaa 4980
gaaattgata caacaatgag atggaaactc ctccctcatt cagatgaacc ttaaaacctc 5040
aaacctagta attttttgtt gatctcctat aaacttatgt tttatgtaat aattaatagt 5100
atgtttaatt ttaaagatca tttaaaatta acatcaggta tattttgtaa atttagttaa 5160
caaatacata aattttaaaa ttattcttcc tctcaacata ggggtgatag c 5211

<210> 37
<211> 5701
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6427133CB1

<400> 37
gctcccaagg ctgagattac tctgcttcat ctggatcgcc catctctggg gtctcatggc 60
tgagtttcag ttccccaatc ctacctgctc ctcaggggcc cagcactggg gctgcaggta 120
ggccacctgt tgagacctgg tgaaagatca ggtataataa tgttctgcag tgaaaagaaa 180
ttgcgtgaag tggaacggat agtgaaagcc aatgaccgtg aatataatga aaagttccag 240
tatgcggata atcgtatcca cacatcgaaa tataatattc tcaccttctt gccaattaat 300
ttatttgaac agttccaaag agtggcaaat gcctattttc tttgccttct gattttacag 360
ctaattccag aaatttcctc cttgacctgg tttaccacca ttgtgccttt ggtcctggtg 420
ataactatga cagctgtcaa agatgccaca gatgactatt ttcgccacaa gagtgataat 480
caagtgaata atcggcagtc tgaagtgctc atcaacagca aactgcagaa tgaaaaatgg 540
atgaatgtca aagtgggaga catcattaaa ttagaaaata accaatttgt tgctgctgat 600
ttacttctcc tatcaagtag tgagccacat ggtctctgtt atgttgaaac tgctgagctt 660
gatggggaaa cgaacctaaa agtccgccat gcactatcag ttacttcaga acttggagca 720
gatatcagca gacttgcagg gtttgatggg attgttgtct gtgaggtgcc taacaacaag 780
ttagataaat tcatgggaat cctttcttgg aaagacagca agcattccct caacaatgag 840
aagataatcc cgagaggctg catcctgaga aataccagct ggtgttttgg aatggttatt 900
tttgcaggtc ctgacactaa actaatgcag aatagtggta agacaaagtt taaaaggaca 960
agcattgata gattgatgaa tactctagta ctatggattt ttgggtttct gatatgcttg 1020
ggaattattc ttgcaatagg aaattcaatc tgggagagtc aaactgggga ccaattcaga 1080
actttcctct tttggaatga aggagagaag agctctgtgt tctccggatt cttaacattc 1140
tggtcatata ttattattct caatacagtt gtacccattt ccttatatgt gagtgtggaa 1200
gtaattcgtc taggacacag atattttata aactgggacc ggaagatgta ttattctcga 1260
aaagcaatac ctgcagtggc tcgaacgacc acgctcaatg aggaactggg gcagattgag 1320
tacattttct ccgacaaaac gggtaccctc actcaaaaca tcatgacctt taaaagatgt 1380
tccattaatg ggagaatcta tggtgaagta catgatgacc tggatcagaa gacagaaata 1440
actcaggaaa aagagcctgt ggatttctca gtcaaatctc aagcggatag agaatttcag 1500
ttctttgacc acaatctgat ggaatccatt aaaatgggtg atcccaaagt tcatgaattc 1560
cttaggttac ttgctctctg ccacactgta atgtcagaag agaatagcgc aggagagctg 1620
atttaccaag ttcagtcacc tgatgaaggg gctctagtga ctgccgctag aaattttggg 1680
ttcatttta aatcccggac cccagagacc ataacaatag aagaattggg aacactagtt 1740
acttatcaat tacttgcctt tttggatttc aacaacacca gaaaaaggat gtctgtcata 1800
gttcgaaacc cagaaggaca gataaagctt tattccaaag gagcagatac tattctgttt 1860
gaaaaacttc atccttccaa tgaagtcctt ttgtctttga cgtcagacca cctcagtgaa 1920
tttgcagggg aaggccttcg gaccttggcc atcgcataca gagacctgga tgacaagtac 1980
tttaaagagt ggcataagat gcttgaagat gcgaatgctg ccacagaaga gagggatgaa 2040
cgaatagctg ggctatatga agaaattgaa agagatttga tgctactagg tgccactgct 2100
gtagaagata agttacagga gggtgttatt gaaacagtta caagtttatc actagccaat 2160
attaagatct gggtcctaac aggagacaaa caagaaactg ccatcaacat cggttatgcc 2220
tgcaacatgc tgactgacga catgaatgat gtgtttgtga tagcagggaa taatgctgtg 2280

gaagtgagag aagaactcag gaaagcaaaa caaaatttgt ttggacaaaa cagaaatttt 2340
tccaatggcc atgtagtttg tgaaaaaaag cagcagctgg agttggattc tattgtagaa 2400
gaaaccataa caggagatta tgccttaatc ataaatggcc acagtttggc tcatgcccta 2460
gaaagtgatg tcaagaatga tctcctagaa cttgcttgca tgtgtaagac tgtaatttgc 2520
tgcagggtca ctccactcca gaaagcccaa gtggtagagc tggtgaagaa gtacagaaat 2580
gctgttactt tggccattgg tgatggagcc aatgatgtca gcatgattaa aagtgctcac 2640
attggtgttg gcatcagcgg ccaggaagga ttgcaagcag tcttagccag cgactattca 2700
tttgcacagt ttagatatct ccaaaggctt ctccttgttc atggaaggtg gtcttatttc 2760
cgaatgtgca aattcttatg ctatttcttc tataagaatt ttgcatttac acttgtgcat 2820
ttctggtttg gtttcttctg tggtttctca gcccagactg tttatgacca gtggttcatc 2880
acccttttta acattgttta cacatcactg cctgttttag ccatggggat ttttgaccag 2940
gatgtgagtg accagaacag cgtggactgt ccccagctct acaaaccagg acagctgaat 3000
ctgcttttta acaagcgtaa atttttcatt tgcgtgttgc atggaatcta cacctcatta 3060
gtccttttct tcatcccctа tggggccttt tacaacgtgg ctggagaaga tgggcaacat 3120
attgctgact accagtcctt tgcagttacc atggccacat ctttggtcat tgtggtcagt 3180
gtgcagatag ccttggatac cagttactgg actttcatta atcacgtctt catctggggg 3240
agcattgcca tttatttctc catttttattt acaatgcaca gtaatggcat ctttggcatc 3300
ttcccaaacc agtttccatt tgttggtaat gcacgacatt ccctgaccca gaagtgcatc 3360
tggcttgtaa ttctcttaac aacagtggct tcagttatgc cagtggtggc attcagattt 3420
ttgaaggtgg atttataccc aaccctgagt gatcagatcc gccggtggca gaaggctcaa 3480
aagaaggcaa ggcctccaag tagccgaagg cctcggaccc gcaggtcaag ctcaagaagg 3540
tctggatatg cttttgctca ccaagaaggc tatggagagc ttatcacatc tggaaaaaat 3600
atgcgagcta aaaatccacc cccaacatca gggctggaaa agacacatta taatagcact 3660
agctggattg aaaatttatg taagaaaacc acagacaccg tgagcagctt tagccaggat 3720
aaaacagtga aactgtgagt caatatgaat ttaaaccacg tagttatctt ttcacttcag 3780
gtggagctga aattctgctg gctccagagt ttgagatttg aggcaagagg tggggcaggc 3840
agattgcctc acttaactta aatctgcggc agacaactgc cagtgcccat caaacaggag 3900
tgtgcgctat ggaaaaccag gccagagggt cactgtctgg tttgtgattt ggtggacaaa 3960
acactcgctg ttacaagtac agatttttttt ttttttaaa tcaacctaga taccaattga 4020
cctgaacttt agaatcttat ttatggagaa aaacttgtaa agctgcatat tcactgaatg 4080
gatcctcagg cggataaaag ggtgcatttt aaaggtatat atccaagctg aaaagcatgc 4140
ctattgacag ataaacatgt atctgtaaga tcagcctttc ccaaggtata cttttaaaat 4200
ttaaagcgtg tactgtgttg ctttcagact gagttgcatg tcactcttta gtcttgatat 4260
ctacctgtct gttcagccag gacaacaaat ggcttccaag cctgaagaat acaaaagtgt 4320
gcttgtgttt ctcatttttа taccagtcta gggacaaagg agactgaaca tctttgcagc 4380
aggataggct ggtaatttga tcaaatttat tcaaaaagct ctcagtctgt gtcatgtaag 4440
gacatgctta tgaaatgtga gagaggctcg ccactaagta ttctaaatac ttttcaatgg 4500
cttttctaac aacctcagta gtaatttgct gagcatcatc cagaccatta atagaatcag 4560
caaagcactg gaatttcaca ctttaatgat aatattccac atagtctatg ggcaaatatt 4620
ttcaacattt ccaatttttа aagcttcaga attgaagcca aacaaattaa taaataattg 4680
ttttaattac tatttaaaaa ctcaggttta gattgtttaa aattagttgc ttttgatact 4740
cagctgtcat gtttataatt caaacatgta gtaaacatat gtaggtaagg ttgtttttтt 4800
ggagatgttg cagctcaaat ttcagtccac atatgaatca tcagtgtatt ttccataaag 4860
tgattcgggc atatttgtgt gaaaacctca gttctgtcac ttcttacctc tataaacttg 4920
gacgataatg tgccttctct gagactcagt ttcttcctct gtaaaatgag gacatactac 4980
ctacctcatg tggttggttg atgattgtct gtcaaagcac aaactctgaa attattaaaa 5040
acataattat ttcataaaca gatgagttaa gttccagtta actcaacatc agtataacag 5100
agcaattgga agagaatatg aaaaaactgg aatctaaata gtcagtgagg aaggctttga 5160
taaaatgaaa ttgccagaaa gatataaaac tggttagggt cctacaggga aataaaatta 5220
taaccgtgga ggtacatttc tctaccagaa agcaaaaata aagcatcatg tcttaatggt 5280
tttctacaaa tcaacttcta attctacaga gtccttaatc tggtccctat taaattcttg 5340
gtcagacaaa gttacatttc ccaagagagt caggtgacac ttgagtgagt ttgatggata 5400
atgagctaat gtgatatcta taggtcacaa ttttttaaaa ccaaaatttt caagtctggg 5460
ataatctttc ctaaatggga tcaaatgaaa taatatgtgt aaaagagtca aatgcagtcc 5520
tttaccatag taactgccta tggacgttgt ctttccctta catgcctgcc tacacttaac 5580
cagatgttgg ttttcaatgt ctaatttgtc attagtttca ccacatttgc tcacttttтg 5640

taacattttt gcaagatttg aaaactttca gtaaatgttt tggcactatt ggtaaaaaaa 5700
a 5701

<210> 38
<211> 1990
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472932CB1

<400> 38
atggctcatg ccccagaacc agacccggcc gccagcgacc tcggggatga gaggcccaag 60
tgggacaaca aggcccagta cctcctgagc tgcatcgggt ttgccgtggg gctggggaac 120
atttggcggt tcccatacct gtgccagacc tatggaggag gtgccttcct catcccctac 180
gtcatcgcgc tggtcttcga ggggatcccc attttccacg tcgagctcgc catcggccag 240
cggctgcgga agggcagcgt cggcgtgtgg acggccatct ccccgtacct cagtggagta 300
ggtctgggct gtgtcacgct gtccttcctg atcagcctgt actacaacac catcgtggcg 360
tgggtgctgt ggtacctcct caactccttc cagcacccgc tgccctggag ctcctgccca 420
ccggacctca acagaacagg ttttgtggag gagtgccagg gcagcagcgc cgtgagctac 480
ttctggtacc ggcagacact gaacatcaca gccgacatca atgacagtgg ctccatccag 540
tggtggctgc tcatctgctt ggcagcctcc tgggcagtcg tgtacatgtg tgtcatcagg 600
ggcattgaga ctacagggaa ggtgatttac ttcacagctt tgttccctta cctggtcctg 660
accatctttc tcatcagagg gctgaccctg ccaggggcaa caaaaggact catctacttg 720
ttcactccca acatgcacat tctccagaac ccccgggtgt ggctggacgc agccacccag 780
atattcttct ctctgtccct ggccttcgga ggacacatcg cttttgcaag ttacaactcg 840
cccaggaatg actgccagaa ggatgcggtg gtcatcgccc tggtcaacag gatgacctcc 900
ctgtacgcgt ccatcgctgt cttctctgtc ctggggttca aagcaactaa tgactgtccc 960
cgcagaaaca tcctcagcct catcaacgac tttgacttcc cagagcagag catctccagg 1020
gacgactacc cagccgtcct catgcacctg aacgccacct ggcccaagag ggtggcccag 1080
ctccccctga aggcctgcct cctggaagac tttctggata agagtgcctc gggcccgggc 1140
ctggccttcg tcgtcttcac ggagaccgac ctccacatgc cgggggctcc tgtgtgggcc 1200
atgctcttct tcgggatgct gttcaccttg gggctatcga ccatgttcgg gaccgtggag 1260
gcggtcatca caccctgct ggacgtgggg gtcctgccta gatgggtccc caaggaggcc 1320
ctgactgggc tggtctgcct ggtctgcttc ctctccgcca cctgcttcac gctgcagtct 1380
gggaactact ggctggagat tttcgacaat tttgccgctt ccctgaacct gctcatgttg 1440
gcctttctcg aggttgtggg tgtcgtttat gtttatggaa tgaaacggtt ctgcgatgac 1500
attgcgtgga tgaccgggag gcggcccagc ccctactggc ggctgacctg gagggtggtc 1560
agtcccctgc tgctgaccat ctttgtggct tacatcatcc tcctgttctg gaagccactg 1620
agatacaagg cctggaaccc caaatacgag ctgttcccct cgcgtcagga gaagctctac 1680
ccgggctggg cgcgcgccgc ctgtgtgctg ctgtccttgc tgcccgtgct gtgggtcccg 1740
gtggccgcgc ttgctcagct gctcacccgg cggaggcgga cgtggaggga cagggacgcg 1800
cgcccagaca cggacatgcg cccggacacg gacacgcgcc cagacacgga catgcgcccg 1860
gacacggaca tgcgctgaag ccggccggag cggggcctgc atgggcgggt ctgtgggggg 1920
gcttggcctg atggtgggcg gggccccgcc cacagggccg accccaatac accagcgact 1980
caaccttgaa 1990

<210> 39
<211> 3760
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8463147CB1

<400> 39
atgacacagg catatcagaa atatattcta gaaaagttac ctaaaagccc tggagacaaa 60
ggcagagcat ggcctgggtc aactccatct gggaatttgc tgtccccatt catggcagct 120
tctaactcct ttcctgagct gtgtagccag gtttccagaa gagagtactg ggacctgcat 180
ggaataccgt ctgaccactt ttctgtgagg gtacaagttg aattctatat gaatgaaaat 240
acatttaaag aaagactaac attatttttc ataacaaacc agagatcaag tctaaggata 300
cgcctgttca atttttctct caaattacta agctgcttat tatacataat ccgagtacta 360
ctagaaaacc cttcacaagg aaatgaatgg tctcatatct tttgggtgaa cagaagtcta 420
cctttgtggg gcttacaggt ttcagtggca ttgataagtc tgtttgaaac aatattactt 480
ggttatctta gttataaggg aaacatctgg gaacagattt tacgaatacc cttcatcttg 540
gaaataatta atgcagttcc cttcattatc tcaatattct ggccttcctt aaggaatcta 600
tttgtcccag tctttctgaa ctgttggctt gccaaacatg ccttggaaaa tatgattaat 660
gatctacaca gagccattca gcgtacacag tgctgcaaat gtgttaatca agttttgatt 720
gtaatatcta cattactatg ccttatcttc acctgcattt gtgggatcca acatctggaa 780
cgaataggaa agaagctgaa tctctttgac tccctttatt tctgcattgt gacgttttct 840
actgtgggct tcggggatgt cactcctgaa acatggtcct ccaagctttt tgtagttgct 900
atgatttgtg ttgctcttgt ggttctaccc atacagtttg aacagctggc ttatttgtgg 960
atggagagac aaaagtcagg aggaaactat agtcgacata gagctcaaac tgaaaagcat 1020
gtcgtcctgt gtgtcagctc actgaagatt gatttactta tggatttttt aaatgaattc 1080
tatgctcatc ctaggctcca ggattattat gtggtgattt tgtgtcctac tgaaatggat 1140
gtacaggttc gaagggtact gcagattcca atgtggtccc aacgagttat ctaccttcaa 1200
ggttcagccc ttaaagatca agacctattg agagcaaaga tggatgacgc tgaggcctgt 1260
tttattctca gtagccgttg tgaagtggat aggacatcat ctgatcacca aacaattttg 1320
agagcatggg ctgtgaaaga ttttgctcca aattgtcctt tgtatgtcca gatattaaag 1380
cctgaaaata aatttcacat caaatttgct gatcatgttg tttgtgaaga agagtttaaa 1440
tacgccatgt tagctttaaa ctgtatatgc ccagcaacat ctacacttat tacactactg 1500
gttcatacct ctagagggca gtgtgtgtgc ctgtgttgca gagaaggcca gcaatcgcca 1560
gaacaatggc agaagatgta cggtagatgc tccgggaatg aagtctacca cattgttttg 1620
gaagaaagta catttttgc tgaatatgaa ggaaagagtt ttacatatgc ctctttccat 1680
gcacacaaaa agtttggcgt ctgcttgatt ggtgttagga gggaggataa taaaaacatt 1740
ttgctgaatc caggtcctcg atacattatg aattctacag acatatgctt ttatattaat 1800
attaccaaag aagagaattc agcatttaaa aaccaagacc agcagagaaa aagcaatgtg 1860
tccaggtcgt tttatcatgg accttccaga ttacctgtac atagcataat tgccagcatg 1920
ggtactgtgg ctatagactt gcaagataca agctgtagat cagcaagtgg ccctaccctg 1980
tctcttccta cagagggaag caaagaaata agaagaccta gcattgctcc tgttttagag 2040
gttgcagata catcatcgat tcaaacatgt gatcttctaa gtgaccaatc agaagatgaa 2100
actacaccag atgaagaaat gtcttcaaac ttagagtatg ctaaaggtta cccaccttat 2160
tctccatata taggaagttc acccacttt tgtcatctcc ttcatgaaaa agtaccattt 2220
tgctgcttaa gattagacaa gagttgccaa cataactact atgaggatgc aaaagcctat 2280
ggattcaaaa ataaactaat tatagttgca gctgaaacag ctggaaatgg attatataac 2340
tttattgttc ctctcagggc atattataga ccaaagaaag aacttaatcc catagtactg 2400
ctattggata acccgccaga tatgcatttt ctggatgcaa tctgttggtt tccaatggtt 2460
tactacatgg tgggctctat tgacaaccta gatgacttac tcaggtgtgg agtgacttt 2520
gctgctaata tggtggttgt ggataaagag agcaccatga gtgccgagga agactacatg 2580
gcagatgcca aaaccattgt gaacgtgcag acactcttca ggttgttttc cagtctcagt 2640
attatcacag agctaactca ccccgccaac atgagattca tgcaattcag agccaaagac 2700
tgttactctc ttgctctttc aaaactggaa aagaaagaac gggagagagg ctctaacttg 2760
gcctttatgt ttcgactgcc ttttgctgct gggagggtgt ttagcatcag tatgttggac 2820
actctgctgt atcagtcatt tgtgaaggat tatatgattt ctatcacgag acttctgttg 2880
ggactggaca ctacaccagg atctgggttt ctttgttcta tgaaaatcac tgcagatgac 2940
ttatggatca gaacttatgc cagactttat cagaagttgt gttcttctac tggagatgtt 3000
cccattggaa tctacaggac tgagtctcag aaacttacta catctgagtc tcaaatatct 3060
atcagtgtag aagagtggga agacaccaaa gactccaaag aacaagggca ccaccgcagc 3120
aaccaccgca actcaacatc cagtgaccag tcggaccatc ccttgctgcg gagaaaaagc 3180
atgcagtggg cccgaagact gagcagaaaa ggcccaaaac actctggtaa aacagctgaa 3240
aaaataaccc agcagcgact gaacctctac aggaggtcag aaagacaaga gcttgctgaa 3300

cttgtgaaaa atagaatgaa acacttgggt ctttctacag tgggatatga tgaaatgaat 3360
gatcatcaaa gtaccctctc ctacatcctg attaacccat ctccagatac cagaatagag 3420
ctgaatgatg ttgtatactt aattcgacca gatccactgg cctaccttcc aaacagtgag 3480
cccagtcgaa gaaacagcat ctgcaatgtc actggtcaag attctcggga ggaaactcaa 3540
ctttgataaa aataaaatga gaaacttttt tcctacaaag accttgcttg aaaccacaaa 3600
agttttgctg gcacgaaaga aactagatgg aaatatatgt aattctctca tatttaaaaa 3660
cgtaatctct tctcttagaa gtatagatca ttttgaaact taatgtacta cttactggta 3720
ctctccctat taatatttga aggacctcaa tggaaagcgg 3760

<210> 40
<211> 1150
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506408CB1

<400> 40
ccagaggaaa ctagtcacaa aaaccctgac tatcacctga tagattgctt gtgctgcctg 60
ataattactc gcacttttcc caggctagtg caaatcttca ggggccgtcc aggactacag 120
agctgtttca ccctaccttg gcttcaatct cttcccccat gctcgaaggt gcggagctgt 180
acttcaacgt ggaccatggc tacctggagg gcctggttcg aggatgcaag gccagcctcc 240
tgacccagca agactatatc aacctggtcc agtgtgagac cctagaagct ccattcttcc 300
aagactgcat gtctgaaaat gctctagatg aactgaatat tgaattgcta cgcaataaac 360
tatacaagtc ttaccttgag gcattctata aattctgtaa gaatcatggt gatgtcacag 420
cagaagttat gtgtcccatt cttgagtttg aggccgacag acgtgctttt atcatcactc 480
ttaactcctt tggcactgaa ttgagcaaag aagaccgaga gaccctctat ccaaccttcg 540
gcaaactcta tcctgagggg ttgcggctgt tggctcaagc agaagacttt gaccagatga 600
agaacgtagc ggatcattac ggagtataca aacctttatt tgaagctgta ggtggcagtg 660
ggggaaagac attggaggac gtgttttacg agcgtgaggt acaaatgaat gtgctggcat 720
tcaacagaca gttccactac ggtgtgtttt atgcatatgt aaagctgaag gaacaggaaa 780
ttagaaatat tgtgtggata gcagaatgta tttcacagag gcatcgaact aaaatcaaca 840
gttacattcc aattttataa cccaagtaag gttctcaaat gtagaaaatt ataaatgtta 900
aaaggaagtt attgaagaaa ataaaagaaa ttatgttata ttatctagac tacacaaaag 960
taagccacac tatatcttca tgagttgcaa atccatggaa acacagtaaa ccagccctga 1020
aacaaagcat ttccttgttt tcagtggtat tagatcttgt ttccacatgt ctgtctcatt 1080
cttcactggg ccttacaggt tagttttaat taactctatg gtatttttct attcttgtct 1140
gatcatgtta 1150

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- [x] BLACK BORDERS
- [x] IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- [x] FADED TEXT OR DRAWING
- [x] BLURED OR ILLEGIBLE TEXT OR DRAWING
- [x] SKEWED/SLANTED IMAGES
- [ ] COLORED OR BLACK AND WHITE PHOTOGRAPHS
- [ ] GRAY SCALE DOCUMENTS
- [ ] LINES OR MARKS ON ORIGINAL DOCUMENT
- [ ] REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- [ ] OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning documents *will not* correct images problems checked, please do not report the problems to the IFW Image Problem Mailbox